CASUALTY ACTUARIAL SOCIETY
FORUM

Winter 2005
Including the Ratemaking Discussion Papers
and Data Management, Quality,
and Technology Call Papers

CASUALTY ACTUARIAL SOCIETY
ORGANIZED 1914

Printed for the Society by
United Book Press
Baltimore, Maryland

NOTICE

The Casualty Actuarial Society *Forum*

Winter 2005 Edition

Including the 2005 Ratemaking Discussion Papers and Data Management, Quality, and Technology Call Papers

To CAS Members:

This is the Winter 2005 Edition of the Casualty Actuarial Society *Forum*. It contains six Ratemaking Discussion Papers and four Data Management, Quality, and Technology Call Papers.

The *Forum* is a nonrefereed journal printed by the Casualty Actuarial Society. The CAS is not responsible for statements or opinions expressed in the papers in this publication. See the Guide for Submission to CAS *Forum* (www.casact.org/aboutcas/forum.htm) for more information.

The CAS *Forum* is edited by the CAS Committee for the Casualty Actuarial Society *Forum*. Members of the committee invite all interested persons to submit papers on topics of interest to the actuarial community. Articles need not be written by a member of the CAS, but the paper's content must be relevant to the interests of the CAS membership. Members of the Committee for the Casualty Actuarial Society *Forum* request that the following procedures be followed when submitting an article for publication in the *Forum*:

1. Authors must submit papers in conformance with the research paper template established by the CAS. See the CAS Web Site at http://www.casact.org/research/template.htm for formatting instructions. Papers not conforming to the template will not be accepted.
2. Authors should submit a camera-ready original paper and two copies.
3. Authors should not number their pages.
4. All exhibits, tables, charts, and graphs should be in original format and camera-ready.
5. Authors should format graphs, tables, exhibits, and text in solid black and white. Avoid using color or gray-shaded graphs, tables, or exhibits as the printing process will produce muddy and moire patterns.
6. Authors should submit an electronic file of their paper using a popular word processing software (e.g., Microsoft Word and WordPerfect) for inclusion on the CAS Web Site.

The CAS *Forum* is printed periodically based on the number of call paper programs and articles submitted. The committee publishes two to four editions during each calendar year.

All comments or questions may be directed to the Committee for the Casualty Actuarial Society *Forum*.

Sincerely,

Glenn M. Walker, CAS *Forum* Chairperson

The 2005 CAS Ratemaking Discussion Papers and the 2005 Data Management, Quality, and Technology Call Papers
Presented at the
2005 CAS Ratemaking Seminar
March 10-11, 2005
New Orleans Marriott
New Orleans, LA

The Winter 2005 Edition of the CAS *Forum* is a cooperative effort between the Committee for the CAS *Forum,* Committee on Ratemaking, and the Committee on Management Data and Information.

The CAS Committee on Ratemaking present for discussion six papers prepared in response to their 2005 call for papers. The CAS Committee on Management Data and Information present four papers for discussion, which were prepared in response to a 2005 call for Data Management, Quality, and Technology papers.

This *Forum* includes papers that will be discussed by the authors at the 2005 CAS Ratemaking Seminar, March 10-11, 2005, in New Orleans, LA.

2005 CAS Ratemaking Discussion Papers

2005 Data Management, Quality, and Technology Call Papers

D&O Reinsurance Pricing - A Financial Market Approach

Athula Alwis, ACAS, MAAA
Vladimir Kremerman, Ph.D.
and Junning Shi, FCAS, MAAA

Abstract
The large number of high severity D&O losses of the past few years has affected the D&O market place creating a serious capacity crunch. The pricing of this line of business has increased dramatically while restricting coverage. This paper will present an objective methodology based on financial market theory to quantify the risk of writing a large D&O reinsurance portfolio. The authors propose that the analysis of the strong correlation between D&O class action law suits and the financial performance of companies is the most critical element in evaluating a D&O portfolio for reinsurance coverage. In addition, the authors will present mechanisms of risk transfer to capital markets based on this new methodology to obtain additional capacity.

Keywords. Class Action Law Suits, Copula, Correlation, Credit Ratings, Credit Spreads, D&O Pricing, Merton Model, Reinsurance, Securities Litigation, Stock Volatility

1. INTRODUCTION

The goal of this paper is to propose an objective pricing methodology based on financial market theory to quantify the risk of writing a public D&O reinsurance portfolio. This paper is not designed to provide a final solution to the very complex problem of underwriting D&O reinsurance. The authors wish to initiate a paradigm shift in the thought process on how to price and structure D&O reinsurance portfolios. It is our belief that the D&O reinsurance must be thought of more as a financial product rather than as an insurance product. The most critical risk that is managed by a D&O policy is the effect of a company's financial performance on its Directors & Officers as well as its shareholders. Therefore, the risk quantification must bear elements of financial risk analysis. In addition, the authors argue that the financial markets represent a natural capacity provider for this cover as long as the risk is quantified in a manner acceptable to financial markets.

The traditional pricing of D&O, both primary and reinsurance, has been largely unsuccessful and at least partly responsible (along with poor risk selection and generous terms & conditions) for the current crisis in the D&O industry. A timely analysis by the Willis Re Professional Liability group in November 2003 indicates that the cumulative cash flow for the industry since 1994 is $0.5 billion and is projected to be negative $13.9 billion for the decade once all incurred claims are paid. During the years 2000 – 2004, the D&O

industry was involved in seven of the largest securities class action settlements of all time. They are as follows:

Rank	Corporation	Settlement Amount
1.	Cendant Corporation	$3.5 billion
2.	Citi Bank	$2.65 billion
3.	Lucent	$517 million
4.	Bank of America	$490 million
5.	Waste Management	$457 million
6.	Daimler/Chrysler	$300 million *
7.	Oxford Health	$300 million

* *There is an on-going second lawsuit by a large investor who did not join the class action law suit settlement in 2003. In addition, there is a third law suit by foreign investors who were excluded from the initial class action law suit.*

The future of the D&O industry looks risky and uncertain to many industry veterans. John Keogh, CEO of National Fire Union (a member of the AIG Group), who provided a more alarming view of the future liabilities, stated that the 57 largest outstanding cases have $966 billion in claimed damages (Learning from Litigation – Interview; Advisen Ltd. 2004). A simple 5.0% to 10.0% settlement range on claimed damages and 50% insurability on losses would indicate a cost of $24 to $48 billion dollars for the industry.

The plaintiffs' law firms have consistently applied innovative methods both in the discovery process and in the actual litigation of class action law suits. The material increase in the amount of settlements has given leading law firms more resources to conduct necessary research in order to pursue new ways to litigate. The Securities Class Action Services (SCAS), a subsidiary of Institutional Shareholder Services (ISS) published the rankings of top plaintiffs' law firms based on securities class actions settlements occurring in 2003. The settlement amounts for the top 7 law firms are as follows:

Rank	Law Firm	Settlement Amount
1.	Milberg Weiss Bershad Hynes & Lerach	$2.1 billion
2.	Bernstein Litowitz Berger & Grossman	$950 million
3.	Grant & Eisenhofer	$611 million
4.	Goodkind Labaton Rudoff & Sucharow	$551 million
5.	Barrack Rodos & Bacine	$390 million

6.	Entwistle & Cappucci	$311 million
7.	Chitwood & Harley	$303 million

The two major regulatory reforms in recent history have not had much effect on the industry. The long term effect of the Private Securities Litigation Reform Act (PSLRA) has been minimal until now. The involvement of institutional investors as lead plaintiff has markedly increased according to a report by Cornerstone Research published in May 2004. In addition, the settlement amounts have been higher when the lead plaintiff is an institutional investor. A National Economic Research Associates (NERA) trend analysis in 2003 indicates that there is no material change in number of filings since the passage of Sarbanes-Oxley Act (SOX). However, the NERA analysis finds a clear decrease in dismissals of law suits. It is clear that both reinsurers and primary carriers should think outside the box in order to quantify and manage this risk if both groups intend to be profitable in the long run.

2. HISTORY AND CURRENT STATUS OF D&O INSURANCE AND REINSURANCE

The history of United States D&O insurance dates back to the 1930s when Lloyd's of London was the main, perhaps only provider of the product. In the 1960s several American insurance companies offered D&O insurance. However, for the most part, Lloyd's underwriting guidelines, claims control procedures, and contract wording were used by the entire industry. In the 60s there were two policies for D&O insurance: a policy covering the corporate reimbursement for indemnification to directors and officers (current Side B); and a policy covering the liability of directors and officers that are not reimbursed by the corporation (current Side A). Eventually, the two policies were combined to form the policy we have today with Sides A and B.

D&O insurance had a profitable run in the 1960s and 70s. However, by late 1970s the claim frequency and severity increased dramatically. In addition, rates decreased and additional coverage was offered due to competition from new entrants into the D&O arena. By mid 1980s, the D&O market was in a severe crisis as several primary companies either markedly reduced the limits or entirely eliminated the product line. Meanwhile, many reinsurers either reduced their capacity or completely left the market. The ensuing hard market in the late 1980s produced significant rate increases, coverage reductions and very specific exclusions. The industry became profitable again. Once again new entrants including captive insurers entered the D&O market and brought in much needed capacity to a profitable segment of the insurance industry. Unfortunately by the early 1990s, the market softening started again. The D&O primary and reinsurance rates were reduced and more coverage was offered. The key turning point in the expansion of coverage was the offering of entity coverage (side C) in mid 1990s for the most part without charging any additional premium to cover the additional risk that was being assumed. The aggregate losses began to rise due to increased frequency and severity. The increase in severity was caused mainly by the shareholder claims based on federal securities laws. However, the continued influx of capacity kept the rates low, limits high and coverage terms and conditions generous for an entire decade.

The high profile financial scandals such as Adelphia, Enron, Tyco International and Worldcom in the last few years were a powerful signal to the D&O market that tough market conditions are inevitable. The hard market was evident in the treaty year 2001 reinsurance renewals as reinsurance capacity was not easily available. Today, both primary and reinsurance prices have materially increased, while nearly 50% of the capacity has left the market. It is clear that the pricing compared to the coverage may have reached the hard market of the early 1990s, however, the cash flow of the industry is expected to be on the negative by billions of dollars once current claims are fully paid.

3. CURRENT D&O REINSURANCE PRICING METHODS AND RELATED ISSUES

Traditional actuarial methods provide experience and exposure rating techniques to price excess reinsurance for D&O policies.

Experience Rating

Experience rating compares primary company developed and trended losses to subject premium adjusted to future rates and exposures. The individual losses are trended for inflation and other influences and then distributed by excess layers. The excess loss development factors are applied to layered and summarized losses. Then, the trended and developed losses are divided by adjusted subject premiums. Various averages are computed in order to obtain the final loss cost. None of these steps is trivial especially in the case of

D&O reinsurance where economic and legal environments change materially in a short time period and thus, history is not a reasonable indicator of the future performance of the portfolio. Below is a partial list of disadvantages in using experience rating of this coverage:

- Change in mix of business in the last several years. There is a shift toward higher attachment points and limits, as well as a change in the mix of risks.
- Change in legal environment and claims consciousness. It is difficult to obtain appropriate loss development factors based on historical experience.
- Trend is affected by economic as well as non-economic factors, such as legal environment, and is not readily available.
- The pricing of high excess layers could be subject to the "free cover" problem. If one intends to overcome this problem with curve fitting, tail adequacy is a complex issue that requires special analysis.
- In both improving and deteriorating underwriting environments, the indications based on experience rating show a material lag.

In general, the approach of looking back at the recent history and pricing a volatile and at times catastrophic product line such as D&O is destined for failure. An indication based on historical experience could not project expected loss costs with reasonable accuracy for the reasons outlined above.

<u>Exposure rating</u>

In current D&O exposure rating, industry data is used to obtain severity distributions. The increased limits factors (ILFs) are computed using these distributions. The amount of time required to gather data and develop necessary severity distributions present an inherent lag in the indications developed using traditional exposure rating. For a volatile product line such as D&O, the lag contained in the traditional exposure rating could produce material uncertainties in the indicated premium need. A partial list of disadvantages related to current exposure rating is presented below:

- The fundamental assumption in the exposure rating that the base pricing being adequate, may not be appropriate for D&O at any point in time. The difference in ILFs is applied to the underlying premium to estimate the reinsurance premium amounts. The adequacy of underlying premium has been highly questionable for a long period of time.
- Due to lack of excess loss data, the credibility of industry severity curves is questionable. The current ILFs used by the practitioners may not reflect the string of class action law suits that the industry faced recently.

- Company specifics might be different from industry, requiring the use of detailed classifications such as private vs. public, large vs. small, IT vs. biotech and so on.
- The industrywide data set may not be rich enough to produce ILFs recognizing differences between various sectors.

4. PROPOSED PRICING METHODOLOGY

The true risk that is covered by D&O insurance is financial risk. Today, more than ever, the trigger for this coverage is linked to the financial performance of the entity. Whenever a public company declares bankruptcy, it is a given that there will be at least one law suit against its directors and officers. In addition, a material drop in stock price while the rest of the sector is performing well or re-statement of previously declared income increases the company's probability of being sued, dramatically. There are other reasons such as misstatements on income or growth, regulatory investigations of accounting and other fraud, SEC investigations on improper activities, prior M&A deals and IPO activity that would spur D&O law suits. It is not clear how the passing of Sarbanes-Oxley (SOX) would affect the frequency of law suits. However, the general expectation by the experienced underwriters is that the Sarbanes-Oxley act would cost the industry more money in the long run. Admittedly, directors and officers of a public company could be sued by a competitor, customer or an employee just as well as by shareholders. It is the frequency and the severity of shareholder law suits that are alarming the whole D&O industry. Therefore, the new pricing methodology is based on the premise that the D&O risk must be quantified and managed as a financial risk.

The insurance pricing theory enables actuaries to estimate averages and standard deviations of a portfolio of risks. The governing theory is the law of large numbers. In financial market theory, the main focus is on risk differentiation. The quality of each risk, the expected loss given default, and the dependencies between each element of risk are individually evaluated and quantified. The portfolio analysis in this process is based on the individual evaluations of each risk.

The new methodology uses credit ratings as a base to establish the expected financial performance of a public company. The credit risk represented by the credit ratings is not directly applied in this methodology. Moreover, the expected financial performance of the public entity is further adjusted using credit spreads, volatility of the stock price and credit spread change and other underwriting adjustments.

The following formula presents the basic methodology:

f(L) = f(M, F, L, C)   where

f(L): Distribution of D&O losses

M: Market Cap of the company

F: Frequency of law suits as a function of default rates, credit spreads, volatility of the stock price and/or credit spreads, regulatory investigations, prior M&A or IPO activity, number of shareholders owning 5.0% or more of the outstanding stock

L: Loss as a function of the market cap

C: Correlation within and between sectors

The goal is to apply this formula to a portfolio of risks simultaneously in a simulation environment and model a distribution of D&O losses that is produced by a correlated multi-variate distribution. The authors use Monte-Carlo simulation to produce the necessary loss distribution because a closed form solution to tackle this problem is not yet developed.

Market Capitalization

The exposure base is the most current market capitalization of the company. It can be argued that the limits are a valid exposure base. The authors argue that market capitalization is a reasonable and perhaps a superior selection as the exposure base. The reasons are as follows:

- It is an independent exposure base that is publicly available and easily verifiable
- It is an objective exposure base that is not dependent on the company management (as opposed to ceded limits)
- There is a reasonable and consistent relationship between the market cap and corresponding losses (refer to Appendix A to see the graph based on an internal Willis analysis)

Number of law suits

The base number of law suits is generated using the publicly available credit ratings such as Moody's and S&P. The fundamental assumption is that each default corresponds to a

D&O law suit. It is possible that some defaults may not trigger class action law suits against directors, however, there are other defaults that would trigger both class action law suits and individual investor law suits. In addition, there are other events such as a drop in stock price that would trigger a class action law suit. The authors expect that class action and investor law suits per given portfolio will exceed the number of credit defaults. Thus, the critical idea is to increase the default rates beyond what is represented by credit ratings in order to capture the frequency of law suits that are above and beyond the frequency of defaults.

There are several critical adjustments that are made to the frequency parameter at this juncture:

- The Moody's and S&P credit ratings are adjusted to reflect the credit outlook of each security and the minimum of the adjusted ratings is selected.
- The credit spreads are used to indicate a credit rating for each company based on default rates indicated by the spreads. Each company's credit rating is further down graded if the credit rating indicated by the spreads is lower than the ratings adjusted for the outlook.
- The volatility of the financial performance is measured using two parameters: volatility of the credit spreads and volatility of the stock price. Based on the volatility index a downgrade of the adjusted credit rating may be recommended.
- If the company is under a regulatory investigation the credit rating has to be adjusted downward to reflect the increased frequency of a law suit.
- If there are institutional investors owning more than 5.0% of the outstanding stock, a downward adjustment of the credit rating is recommended.
- If there has been any M&A activity or an Initial Public Offering during the past three years by the company, a downward adjustment of the credit rating is recommended.

The downgraded credit ratings replace the original credit ratings for the companies in the portfolio prior to simulation. A mathematical model based on this approach requires a thorough calibration process to determine the appropriate level of downward adjustments that are necessary to produce an appropriate number of law suits during the simulation process.

<u>Loss as a function of the Market Cap</u>

The historical relationship between losses from law suits and the market capitalizations prior to the law suits were examined using a database containing about 1200 cases. The results are presented in Appendix A (Figure 1). The loss as a percentage of market capitalization seems to decrease at a decreasing rate as the market capitalization increases. It is beyond the scope of this paper to further analyze the shape of this curve. Given sufficient data, it would be a useful exercise to learn at what threshold (if any), the size of loss as a percentage of market capitalization begins to increase. The volatility around these severity numbers is recognized during the simulation by introducing a random distribution. There are several examples of very large settlements (Cendant above $3 billion and Citi Bank at $2.65 billion) that warrant a material variation around average loss severities by market cap during the simulation process.

<u>Correlation within and between sectors</u>

It is clear from the past experience that the D&O law suits are correlated. For example, the explosion of law suits in the IT industry due to IPO laddering or accounting scandals in many high flying public companies make it evident that the correlation among law suits exists and it is material. It is also possible that large law firms do research on industries and sectors as a whole when long term strategies to bring law suits are designed and planned. In our analysis, we have projected material correlation within industry sectors and a nominal amount of correlation between sectors. It is extremely important to recognize the potential for correlated loss events when generating aggregate D&O losses. As stated at the beginning, our goal is to quantify the risk of writing a portfolio of D&O losses as a reinsurer. It is important to know the average loss so that the basic pricing can be completed. However, it is more important to know the variability around the average loss because reinsurance is bought for the most part to control that variability. If correlation assumptions are not included in the analysis, the tail of the loss distribution would not reflect the true nature of this risk (i.e. the tail would be too weak to predict a reasonable range of future expected losses.) In addition, if capital is allocated as a function of the 99th percentile loss of the loss distribution, then the loss distribution must reflect correlated events to truly reflect the size and the probability of a very large aggregate loss. The technical aspects of the building of a correlation matrix are presented in appendix C where the authors attempt to obtain defaults in a correlated multivariate environment.

5. REINSURANCE AND CAPITAL MARKET PRODUCTS

The output from the modeling described in Section 4 is a distribution of D&O losses that could be used to structure and price the portfolio, to understand the effects of aggregations, and to allocate capital based on parameters determined by the analyst and the client. The authors do not intend to discuss issues relating to the parameter risk involving the development of the loss distribution proposed in this paper, but would like to caution practitioners to be aware of its presence in this type of financial modeling.

Quota Share Reinsurance

It is a fairly simple process to model commissions and other expenses once the distribution of losses is determined to a certain level of credibility. The calculation of loss ratios and combined ratios would follow. The developed loss distribution allows the analyst to estimate not only the mean and standard deviation of the D&O portfolio but also gives an opportunity to estimate higher moments (skewness and kurtosis).

If the capital allocation assumptions are agreed upon (for example, requiring capital to cover the unexpected portion of a 1 in 100 year loss), then the return on allocated/indicated capital calculation is a straight-forward process based on the developed loss distribution.

Excess of Loss Reinsurance

The gross distribution developed from simulation should be layered (per name/per account) according to the limits, aggregate limits, retentions and other conditions of the reinsurance contract to obtain the excess of loss distribution. Then, the pricing of the excess layer and the development of risk/reward measures for the reinsurance transaction becomes a straightforward exercise. In addition, the understanding of the aggregate losses within layers is a valuable insight in both pricing and risk management since aggregate limits and reinstatement premiums can be computed in an efficient manner. Exhibit 1 contains a sample D&O portfolio and exhibit 2 contains the output from modeling exercise.

The calculation of return on indicated capital becomes a routine procedure due to the availability of both gross and net loss distributions. It is an interesting exercise to see the changes in loss costs, indicated capital and return on capital as certain names (accounts) are added and subtracted from the portfolio. The allocation of capital to the portfolio based on marginal cost of capital needed to write the risk is a reasonable and appropriate methodology. An efficient portfolio optimization tool that uses the above developed distribution could add tremendous value to the entire process.

Stop Loss Cover (Structured similar to a CDO)

The reinsurance portfolio of the primary D&O carrier should be constructed by offering pre-determined limits such as $25 million up to $100 million to its clients. By limiting the number of accounts in the portfolio to about 200 names, one can construct a portfolio around $10 billion. Then, the portfolio can be structured in tranches (of say, $200 million) to be sold to reinsurers, hedge funds and other investors. The primary carrier should retain, for instance, the first tranche of $200 million thus, increasing the quality of risk in higher tranches. The cost for higher tranches should decrease materially as reinsurers, hedge funds and investors are further removed from the risk of a loss. The analytical method that is outlined in the paper lends to determining the quality of the risk, the variation around the mean and various percentiles for specified tranches. By developing a distribution of aggregate losses in the way proposed in this paper, primary insurance companies will be able to present an objective and an independent methodology to quantify the risk of writing this type of cover.

Potential Future Development: Option Pricing based on the Wang Transform

The "Wang Transform" introduced by Shaun Wang (2002) can be applied here using the probabilities derived from the aggregate distribution and estimating the "Market Price of Risk" (a.k.a. Sharpe Ratio (λ)) based on the underlying market data of the companies in the portfolio. One clear difference in this methodology compared to Shawn Wang's methodology for insurance risks is that one can compute the Market Price of Risk based on the underlying data by following the approach proposed in this paper. This is a material advantage of treating D&O as a financial product as opposed to an insurance product.

6. CONCLUSION

We have presented an objective and independent methodology to quantify the risk of writing a large public D&O reinsurance portfolio. This is a starting point rather than a final solution to a very complex problem. It is our sincere hope that with our work, we have started a paradigm shift in the thought process on how to assess risk vs. reward in writing D&O reinsurance. Please note that any model, however sophisticated, will not replace good old fashioned underwriting required prior to and during the risk selection process. In the final analysis, this methodology and future reinsurance pricing models based on the methodology must be viewed as tools designed to enhance the total underwriting process.

Appendix A: Model and Related Adjustments

Our D&O model consists of 2 main parts.

1. The adjustments to the initial credit ratings (presented in Section 4)
2. The simulation engine containing final ratings, severity curves and the correlation matrix

Comments on adjustments to credit ratings:

- The adjustment based on the credit spreads is predicated on the formula derived from the "Reduced Form Approach" by Lubochinsky (2002). The formula is as follows:

  Spread (S) = $d*(1 – R)*(1+r)/[1- d*(1 – R)]$, where d: indicated default rate; R: recovery rate; r: risk free rate

The value for d represents the new adjusted credit rating for the security.

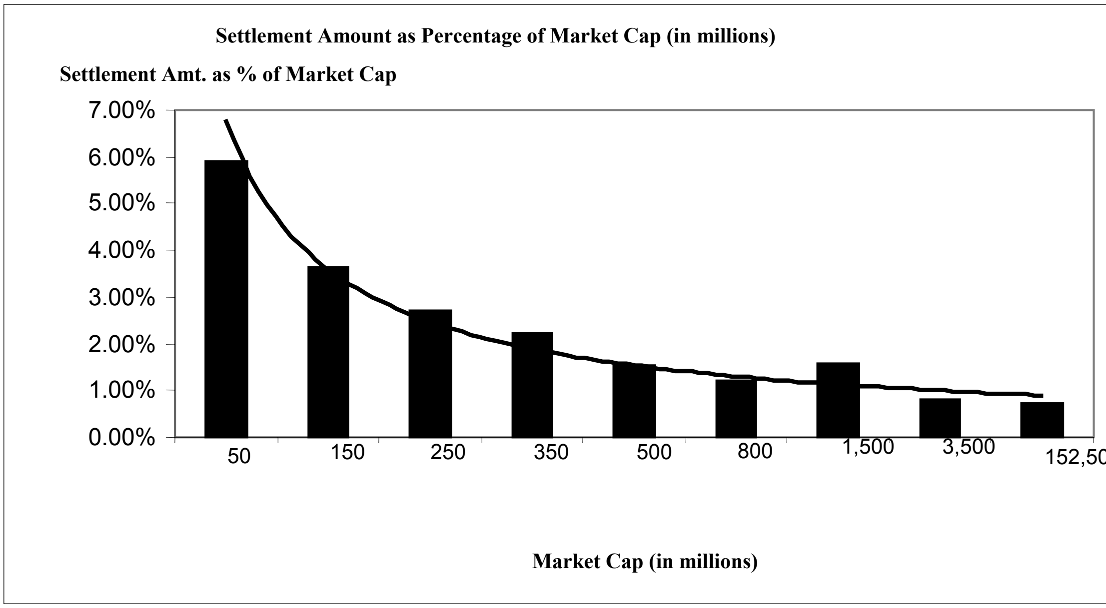
- The most difficult adjustment is to determine the necessary down grades based on stock and credit spread volatility. There are many securities litigation suits that are based on sudden drops in stock price, income and growth not matching the stated stable numbers predicted by the management and re-statement of income due to poor financial performance as well as outright fraud. Therefore, it is extremely important to capture the volatility of the stock price as a predictor of future law suits.

  1. Compare the β (volatility of the stock) to the β for the industry sector. For example, the volatility of a technology company stock should be measured against the rest of the technology sector not against the general market. The model contains confidence interval that set the downgrades to one, two or three notches.
  2. The volatility of the credit spreads is compared to the average movement of the spreads for the industry. The proprietary confidence intervals determine the extent of the downgrades to the previously adjusted credit ratings.

The maximum of the two sets of downgrades is selected as the adjustment for this step.

The chart below (courtesy of Thomson Financial) presents a comparison of the movement of IBM's stock price against industry indices.

Willis Analytics Figure 1



Appendix B: Credit Spreads

The credit ratings published by the ratings agencies provide a measure of financial health of a public company. However as demonstrated in cases such as Enron, there is a certain time interval between change of the company's financial situation and change of its credit rating. Credit spreads (the difference between yields on risk free bonds and corporate bonds) are available instantaneously and reflect bad news well ahead of credit ratings. According to Schonbucher (2003), credit spreads contain the market's opinion on the default risk of the obligor. They provide an objective, market based early warning instrument for changes in the default risk of the obligors. Thus, credit spreads, though volatile, provide a more timely measure of a company's debt paying ability, hence, financial health. There is an intense discussion in current literature regarding the kind of information that is embedded in spreads. Customarily, value of a spread is expressed as a sum of default spread and residual spread, (Lubochinsky (2002)). The first component, default spread, is a direct measure of risk of default. According to the "reduced form" approach (Lubochinsky (2002)), the default spread is proportional to the risk of default without recovery. However, default spread may not always be the main component of the spread. For example, it is shown that for AAA entities only 5% of credit spreads is attributable to risk of default (see Delianedes, Geske (2001)). The residual spread is influenced by taxes, jumps, liquidity, and market risk factors.

Appendix C: Correlation

A simple definition of linear correlation:

Correlation is the degree to which two or more quantities are linearly associated. In a two-dimensional plot, the degree of correlation between the values on the two axes is quantified by the so-called correlation coefficient.

According to Li (1999): the linear correlation of default for two securities *i* and *j*, ρ_{ij} satisfies the following equation

$$\rho_{ij} = \frac{Cov(i,j)}{\sqrt{Var(i)\cdot Var(j)}} = \frac{Cov(i,j)}{\sqrt{u_i(1-u_i)u_j(1-u_j)}} \qquad (1),$$

where u_i, u_j are corresponding default probabilities. The approach the authors used to incorporate correlation into the simulation engine is described below. Any attempt to simulate credit events without giving appropriate regard to the effects of correlation would severely underestimate the tail of the distribution.

How to build a correlation matrix for simulation

It is necessary to compute $Cov(i,j)$ using the within and between correlations assumptions determined at the outset of the analysis. The authors use Merton's approach to calculate $Cov(i,j)$. The Merton approach to the firm's value suggests that a default occurs when the value of assets is below certain threshold (Merton (1974)). In other words, default takes place when a random variable representing firm's assets X_i (with CDF $P(X_i)$) is below a certain level. Two companies are in default if $X_i < P^{-1}(u_i),\ X_j < P^{-1}(u_j)$. Then the covariance equals to

$$P(P^{-1}(u_i), P^{-1}(u_j)) - u_i u_j$$

In order to generate credit events (defaults), Monte Carlo simulation is applied. A set of independent normal random variables are transformed to a correlated standard normal random variables by introducing a correlation matrix to the process. The correlated standard normal random variables are compared to the thresholds based on default rates. The correlation matrix needs to be decomposed based on the Cholesky decomposition prior to creating a matrix of correlated standard random variables. There are two key adjustments that are necessary to obtain a reasonable set of outcomes. They are outlined below.

<u>The Merton Adjustment</u>

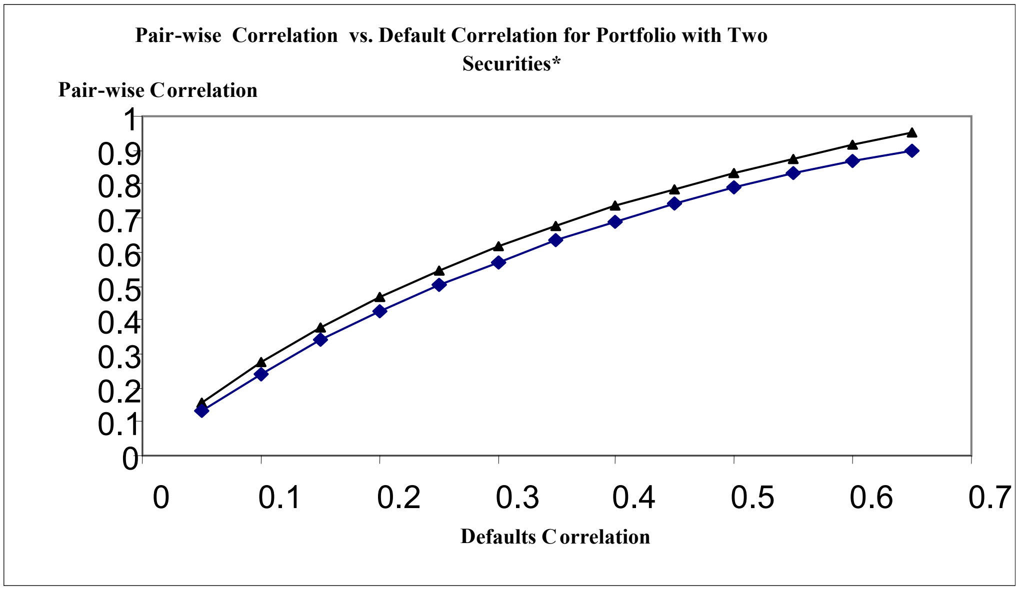
The most straightforward approach in the calculation of a copula in the above equation is to assume that X_i, X_j are normally distributed. Then, one obtains for coefficient of correlation (Pugachevsky (2002))

$$\rho_{ij} = \frac{N^{(2)}(N^{-1}(u_i), N^{-1}(u_j), \rho_{ij}^M) - u_i u_j}{\sqrt{u_i(1-u_i)u_j(1-u_j)}} \qquad (2).$$

In equation (2), $N^{(2)}$ is the cumulative bivariate normal distribution function with pair-wise correlation coefficient ρ_{ij}^M and $N^{(-1)}$ is the inverse of standard normal distribution. The matrix ρ_{ij}^M is determined numerically from Eq.(2) and is used in the loss simulation. Figure 2 contains graphs of ρ_{ij}^M for two arbitrarily selected probabilities of default as a function of events correlation ρ_{ij}. After pair-wise correlation coefficients are computed, the simulation engine can produce a correlated multi-variate distribution. According to Pugachevsky (2002), the main advantage of this method is that it is easy to define correlations between random variables in a simulation environment. The fact that this method does not project the time of default is a weakness in general, however, it is not a major issue for the D&O reinsurance pricing methodology that the authors propose. Schonbucher (2003) presents a model using a generalization of the Archimedean Copula as opposed to the Normal copula that is applied in this model to capture the timing of default.

It is evident from Fig. 2 that the resulting correlation ρ_{ij}^M is materially higher than the discrete events correlation ρ_{ij}. It should be noted that the use of events correlation ρ_{ij} in simulation would lead to substantial underestimation of the correlation effect.

Willis Analytics Figure 2



*Default probabilities are (0.1, 0.05) and (0.05, 0.05) for upper and lower lines, respectively.

<u>How to make the correlation matrix positive-definite</u>

The resulting correlation matrix obtained from Eq. (2) is not necessarily positive-definite. The positive-definiteness is a requirement that guarantees the ability to decompose the correlation matrix after the application of Merton adjustment. There are several known techniques that would help transform the correlation matrix into a positive definite matrix. The authors chose the approach suggested by Rebonato and Jackel (1999) to revise the matrix ρ_{ij}^{M}. The adjustment procedure involves three steps. First, eigenvalues and eigenvectors of the matrix of pair-wise correlations Σ^2 are defined,

$$\Sigma^2 S = \Lambda S ,$$

where Λ, S are matrices of eigenvalues and eigenvectors, respectively. Second, zero or negative eigenvalues are replaced by very small positive numbers. Third step involves the production of the correlation matrix using modified eigenvalues λ' and eigenvectors of initial correlation matrix. Taking into account that diagonal elements of the correlation matrix have to be equal to one, the resulting modified matrix equals

$$\sqrt{T} S \Lambda' S^T \sqrt{T'}$$

where the matrix T is required for the normalization.

Acknowledgements

The authors would like to thank William Panning, Managing Director, Willis Analytics and Sean Whelan, Practice Leader, Willis Re Professional Liability for their continued support, direction and valuable comments. The authors would also like thank Michael McManus, Senior Vice President and Actuary, Chubb Group of Insurance Companies, Yves Provencher, Lixin Zeng, Peter Szendro, Gene Mason and Kelly Edmunds, Senior Vice Presidents at Willis Re and Mark Hvidsten, Managing Director, Willis Re London, for reviewing our initial draft and providing constructive comments. Finally, the authors would like to thank George Reeth, Jr. CEO, Willis Re North America and his senior management team for providing the opportunities and resources for long term research and development.

The authors appreciate the constructive feedback from the CAS Rakemaking Committee members Beth Fitzgerald, Jonathan White, Eric Johnson and John Pedrick. In addition, the technical assistance from Jason Harger to explore financial databases is much appreciated.

Bibliography

1. Aronowitz, D. (2003): The D&O Market – One Year After the White Paper, Professional Liability Underwriting Society Journal
2. Bajaj, M., Mazumdar S. C., Sarin, A. (2000). Securities Class Action Settlements: An Empirical Analysis, Haas School of Business, University of California-Berkeley and Leavey School of Business, Santa Clara University
3. Beaver, W. H., Malernee, J. K. (1990): Estimating Damages in Securities Fraud Cases, Cornerstone Research
4. Bohn, J. R. (1999): Characterizing Credit Spreads, Haas School of Business, University of California
5. Boyer, M. (2003). Directors and Officers Insurance and Shareholders Protection
6. Buckberg, E., Foster, T., Plancich, S. (2004): Recent Trends in Securities Class Action Litigation: 2003 Early Update, NERA Economic Consulting
7. Das, S. R., Freed, L., Geng, G., Kapadia, N. (2002): Correlated Default Risk, Leavey School of Business, Santa Clara University
8. Delianedis, G. and Geske, R. (2001) : The component of corporate credit spreads: Default, Recovery, Tax, Jumps, Liquidity, and Market Factors", eScholarship Repository, University of California
9. Duffie, D., Singleton, K. J. (1999): Modeling Term Structures of Defaultable Bonds, The Review of Financial Studies, Oxford University Press
10. Gische, D. M. and Fishman V.: Directors and Officers Liability Insurance – Overview
11. Glasserman, P. (2003): Tail Approximations for Portfolio Credit Risk, Columbia Business School
12. Li, Davis X. (1999): On Default Correlation: A Copula Function Approach. The RiskMetrics Group, Working Paper Number 99-07
13. Lockwood, G. (2003): D&O Insurance: The Cycle of Change, Professional Liability Underwriting Society Journal
14. Lubochinsky, C. (2002): "How much credit should be given to credit spreads", Financial Stability Review
15. Merton, R. C. (1974): On Pricing of Corporate Debt: The Risk Structure of Interest Rates, Journal of Finance, 29, pp. 449-470
16. Mina, J. (2001): Mark-to-Market, Oversight, and Sensitivity Analysis of CDO's, Risk Metrics Group
17. Moody's Investor Services (2003): Default & Recovery Rates for Corporate Bond Issuers; A statistical review of Moody's Ratings Performance, 1920-2002
18. Pugachevsky, D. (2002): Correlations in Multi-Credit Models. 5th Columbia-JAFEE Conference on Mathematics in Finance, 5-6 April 2002
19. Rebonato, R., Jackel, P. (1999): The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes. Quantitative Research Centre of the NatWest Group.

20. Rogge, E., Schonbucher, P. J. (2003): Modeling Dynamic Portfolio Credit Risk, Department of Mathematics, Imperial College and ABN AMRO Bank, London and Department of Mathematics, ETH Zurich, Zurich
21. Simmons, L. E. (1999): Securities Law Suits: Settlements Statistics for Post-Reform Act Cases, Cornerstone Research
22. Simmons, L. E., Ryan, E. M. (2004): Post-Reform Act Securities Law Suits: Settlements Reported Through December 2003, Cornerstone Research
23. Standard & Poor's Risk Solutions (2003): Corporate Defaults Peak in 2002 Amid Record Amounts of Defaults and Declining Credit Quality – Hazards Remain
24. Szendro, P. (2003): Members of the Board – Will Your D&O Policy Defend You in Bankruptcy?, Professional Liability Underwriting Society Journal
25. Wang, S. S. (2002): A Universal Framework for Pricing Financial and Insurance Risks, ASTIN Bulletin

Glossary

Call Option An option contract giving the owner the right (but not the obligation) to buy a specified amount of an underlying security at a specified price within a specified time.

Collateralized Debt Obligation (CDO) is an investment-grade security backed by a pool of bonds, loans, and other assets. Investors bear the credit risk of the collateral. Multiple tranches of securities are issued by the CDO, offering investors various maturity and credit risk characteristics

Copula A function that joins univariate distribution functions to form multivariate distribution functions.

Credit Risk is the risk due to uncertainty in a counterparty's (also called an obligor or credit's) ability to meet it's obligations. Because there are many types of counterparties, from individuals to sovereign governments and many different types of obligations, from auto loans to derivatives transactions, credit risk takes many forms

Credit Spread for a bond equals to difference between yield on a risky bond and yield on a default-free government bond with a similar maturity

Market Capitalization (Market Cap) is the total dollar value of all outstanding shares

Private Securities Litigation Reform Act (PSLRA) was enacted by the Congress in 1995 to discourage "meritless" securities class action litigation. The Act introduced a "Hightened Pleading Standard" requiring plaintiff to "state with particularity the facts giving rise to a strong inference that the defendant acted with the required state of mind". Automatic Stay of Discovery provides that discovery is stayed when a defendant files a dispositive motion to dismiss in a securities fraud claim. That allows defendant not to produce any documents that trial lawyers have demanded while the court decides the motion to dismiss

Recovery Rate In the event of a default, the recovery rate is the fraction of the exposure that may be recovered through bankruptcy proceedings or some other form of settlement

Sarbanes-Oxley (SOX) Act was passed by Congress in 2002 to "protect investors by improving the accuracy and reliability of corporate disclosures made pursuant to the securities laws, and for other purposes"

Sharpe Ratio (Market Price of Risk) is the difference between the return on a security and the return on a benchmark portfolio divided by the standard deviation of the return on the security; differential return per unit of risk

Biographies of Authors

Athula Alwis is Vice President, Analytical Services at Willis Re Inc in New York City, New York. He is responsible for structuring, pricing and capital modeling of large reinsurance transactions in the financial products arena. Athula has a BS in Mathematics from University of Colombo, Sri Lanka and a MS in Mathematics from Syracuse University, New York. He is an Associate of the Casualty Actuarial Society (CAS) and a member of the American Academy of Actuaries (AAA). Athula is a member of the AAA Risk Based Capital committee, and is a frequent presenter at industry conferences. Athula co-authored a research paper titled "Credit & Surety Pricing and the Effects of Financial Market Convergence" in 2002.

Vladimir Kremerman is Junior Actuary, Analytical Services at Willis Re Inc. in New York City, New York. He is responsible for property/casualty reinsurance analysis. Vladimir has a Ph.D. in Physics from Vilnius State University. He worked as a physicist at Semiconductor Physics Institute of Lithuanian Academy of Sciences, and Center for Ultrafast Photonics in City University, New York. He authored/coauthored several papers on statistical mechanics.

Junning Shi is Senior Vice President, Analytical Services at Willis Re Inc. in New York City, New York. He provides actuarial consulting services in specialty lines such as marine, aviation, surety, international & retro as well as in many other property & casualty lines. Junning has a Doctor of Arts in Mathematics from Idaho State University. He is a Fellow of the Casualty Actuarial Society (CAS) and a member of the American Academy of Actuaries (AAA). He has authored/co-authored several papers on approximation theory.

Willis Analytics **Exhibit 1**
D&O Reinsurance Pricing

Test Portfolio

Index	Account Name	Market Cap	Sector	Original Rating	Adjusted Rating	Loss as % of Mkt Cap	Std. Dev.
1	Company 1	5,615,101,390	6	A2	Baa1	0.73%	5.00%
2	Company 2	1,247,762,880	3	Baa2	Baa3	1.59%	10.00%
3	Company 3	221,642,688	4	B1	B3	2.73%	10.00%
4	Company 4	210,080,000	1	Ba3	B1	2.73%	10.00%
5	Company 5	196,820,000	7	A3	Baa1	3.64%	10.00%
6	Company 6	166,790,000	4	Ba2	B2	3.64%	10.00%
7	Company 7	162,630,000	8	Aaa	Aa1	3.64%	7.00%
8	Company 8	161,460,000	9	Baa1	Baa2	3.64%	9.00%
9	Company 9	156,520,000	10	A3	Baa1	3.64%	10.00%
10	Company 10	149,890,000	11	A3	Baa2	3.64%	15.00%
11	Company 11	148,200,000	2	B1	B2	3.64%	15.00%
12	Company 12	144,560,000	5	B1	B2	3.64%	15.00%
13	Company 13	136,890,000	1	Caa1	Caa3	3.64%	15.00%
14	Company 14	126,620,000	5	Baa3	Ba1	3.64%	15.00%
15	Company 15	112,710,000	12	Baa1	Baa2	3.64%	15.00%
16	Company 16	108,550,000	13	Aaa	Aa1	3.64%	15.00%
17	Company 17	104,910,000	3	Ba1	B3	3.64%	15.00%
18	Company 18	98,930,000	1	Ba2	Ba3	5.91%	15.00%
19	Company 19	95,680,000	4	Ba3	B1	5.91%	15.00%
20	Company 20	93,340,000	3	A1	A3	5.91%	15.00%

Willis Analytics
D&O Reinsurance Pricing

Exhibit 2

Test Portfolio

Reinsurance Terms	LAYER 1	LAYER 2	LAYER 3	LAYER 4
Per Risk Limit	999,999,999,999	2,000,000	3,000,000	15,000,000
Per Risk Attachment	-	-	2,000,000	5,000,000
Aggregate Limit	999,999,999,999	999,999,999,999	9,000,000	30,000,000
Aggregate Deductible	-	-	-	-

Percentiles	LAYER 1 Losses	LAYER 2 Losses	LAYER 3 Losses	LAYER 4 Losses	LAYER 1 Counts	LAYER 2 Counts	LAYER 3 Counts	LAYER 4 Counts
Mean	4,351,446	1,197,609	1,149,676	1,616,123	0.73	0.73	0.49	0.29
Std Dev	11,230,421	2,291,762	2,447,945	4,665,260	1.35	1.35	1.00	0.69
C.V.	258%	191%	213%	289%	186%	186%	203%	238%
Median	0	0	0	0	0	0	0	0
Min	0	0	0	0	0	0	0	0
Max	221,944,867	27,346,732	32,941,040	59,328,614	16	16	12	10
65.0%	974,793	974,793	-	-	3	3	0	0
70.0%	2,578,211	2,000,000	353,564	-	3	3	1	0
75.0%	6,376,881	2,000,000	3,000,000	712,352	4	4	1	1
80.0%	13,006,533	2,000,410	3,000,000	6,921,037	4	4	1	1
85.0%	23,952,844	2,079,845	3,000,000	15,000,000	5	5	1	1
90.0%	45,372,773	2,960,912	3,000,000	15,000,000	6	6	1	1
95.0%	82,476,651	4,007,874	6,000,000	15,000,000	7	7	2	2
96.0%	94,504,038	4,104,277	6,000,000	16,705,200	8	8	2	2
97.0%	110,950,420	4,532,038	6,000,000	21,781,801	8	8	2	2
98.0%	129,181,731	5,729,388	6,000,000	30,000,000	9	9	2	2
99.0%	157,752,527	6,062,463	8,232,189	30,000,000	10	10	3	2

Ratemaking for Captives and Alternative Market Vehicles

Ann M. Conway, FCAS, MAAA

Abstract:
Although captives represent a significant part of the insurance market, there is relatively little information on the subject in the actuarial literature. This paper describes ratemaking techniques and approaches that can be used for captives and other alternative market vehicles. To put the discussion in context, the paper begins with a description of various captive structures. It conclude with a discussion of the financial considerations associated with a captive program.
Keywords: Ratemaking, Expense Loads, Increased Limits, Trend and Loss Development, Solvency, Credibility, Simulation.

1. INTRODUCTION

Although captives have been in existence since the nineteenth century (the protection and indemnity clubs used for marine exposures essentially functioned like group captives (defined later)), the use of these vehicles has grown significantly in the last 40 years. This growth has been driven, in part, by the high cost or lack of availability of commercial insurance (i.e., in a hard market situation). Over time, captive owners often find reasons beyond the market cycle to use and/or expand their captives; in recent years, employee benefits have become an area of interest for U.S. concerns.

In spite of the prevalence of these alternative market vehicles, there is relatively little information on the subject in the CAS literature. For example, there is only one reference to captives in the CAS search database (Reinsuring the Captive/Specialty Company, Lee Steeneck, CAS 1982) and there are a limited number of references to self-insurance (e.g., Statistical and Financial Aspects of Self-Insurance Funding, Leigh Halliwell, CAS 1996; Hospital Self-Insurance Funding: A Monte Carlo Approach, David Bickerstaff, CAS 1989; Simulation Models for Self-Insurance, Trent Vaughn, CAS 1996). The self-insurance papers in the CAS literature focus mainly on simulation techniques.

This paper will describe ratemaking techniques and approaches that can be used for captives and other alternative market vehicles. Issues addressed include:

- Understanding the data submission – There is a wide range in the types of information that are compiled in the initial stages of evaluating the feasibility of a captive. This paper will discuss ways to evaluate the data provided and how to identify potential problems, omissions and inaccuracies.

- Evaluating the exposure – In the hard market captives are taking on new and different exposures than what they traditionally assumed. This paper will discuss potential exposures and how to evaluate them for ratemaking purposes. In some cases the exposures can be evaluated using more traditional techniques (e.g., increased limits factors) but in other cases no traditional methods exist (e.g., coverage for the workers compensation liabilities associated with smallpox vaccination).
- Making rates with limited or no data – The paper will discuss methodologies and approaches that can be used to address the limitations in the available data and describe how other data (e.g., insurance industry information or other types of data) can be incorporated into the pricing model. It will also discuss how the approach can evolve over time to incorporate more of the captive's data. The paper will also discuss how traditional ratemaking techniques (e.g., increased limits factors, credibility) can be applied in the context of ratemaking for captives and other alternative market entities.
- Reviewing allocation models – An end-product of many captive rate analyses is rates by entity or cost center. This paper will briefly describe some of the more commonly used allocation approaches and evaluate their strengths/weaknesses.
- Identifying some of the common pitfalls encountered in these types of pricing analyses.

The paper will begin with a description of the various alternative market structures to put the remaining points in context.

2. CAPTIVE BASICS

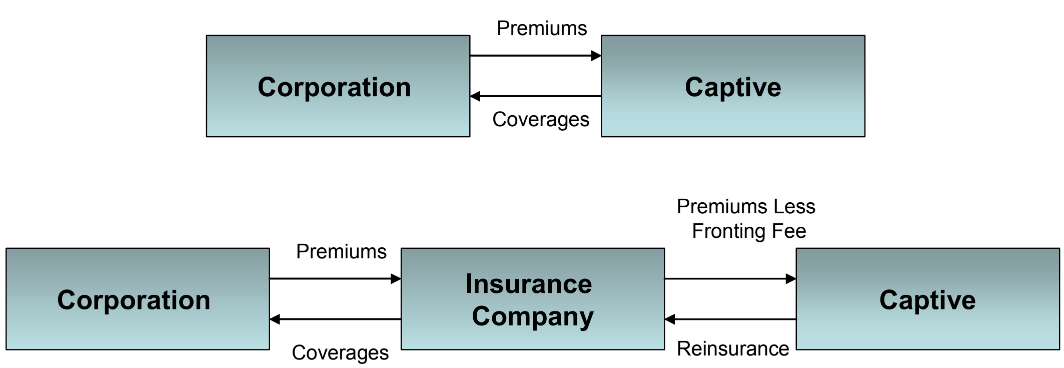
According to Best's Captive Directory, a captive can be defined as a closely held insurance company, where much or all of the captive's business is typically supplied by and controlled by its owners. Typically the owners (or shareholder/insureds) are actively involved in the underwriting, operations and investments of the captive. As of December 31, 2003 there are over 5,000 captives licensed, both in the U.S. and overseas, not including "cell" captives, according to the Best's Captive Directory. (Cell captives will be defined below.)

2.1 Types of Captives

There are a number of potential captive structures, including:

- Single parent – In this case, a single parent owns the captive (and a parent may own more than one captive) and the captive's financial results are "rolled up" to the parent. There are two types of single parent captives:
  - Direct writing – Under this structure, the captive issues the policy or policies directly to its insured (its parent). This type of captive is often used for coverages where the parent typically would not need to provide certificates of insurance or where there are fewer regulatory constraints (i.e., excess coverages, deductible reimbursement policies, or indemnification policies for liability).
  - Fronted or reinsurance captive – A fronting insurance company (a front) issues the policy to the parent and then the captive reinsures some or all of the front's exposure. A fronted program is often used for primary coverages where there is a need to issue insurance certificates regularly (i.e., for workers compensation). Fronting adds an additional level of expense to a captive operation because the front needs to be compensated for the use of its name, the administrative expenses of issuing the policy, and its creditworthiness or Best Rating. Currently fronting fees run between 10 and 12% of premium; in a softer market those fees could be lower. On a reinsurance basis, a captive can also participate in the excess layers provided by the captive's reinsurers or in the excess coverage purchased by the parent. In this scenario, there would not be a fronting fee.

    The following chart shows the relationships between a corporation and its captive under both a direct and fronted program.



Given the cost of establishing and maintaining a captive, the single parent option is not always viable for smaller entities. An often used "threshold" premium for a single parent captive is $750,000 to $1,000,000, given the level of ongoing captive expenses. Many organizations, however, establish single parent captives at significantly lower premium levels.

- Group captive - A group captive is owned by a group of companies; these may be industry specific (i.e., a group of nursing homes, a consortium of educational institutions) or may cross industry segments. Industry specific captives are often referred to as association captives, a name derived from the fact that many such captives are formed by members of a trade association. These captives can design their program around the particular exposures of their participants; the trade-off to the homogeneity of their exposure is that it may lack the risk spread that a more diversified captive would provide. Group captives can either write directly or on a reinsured basis (i.e., the captive reinsures the front). Results are shared among the captive owners in accordance with the participation agreement. Group captives often require members to contribute capital when they join the program; this contribution is often set as a percentage of direct written premiums. Note that some group captives write unrelated business in addition to the exposures of the group owners. While this may support the tax deductibility of premium (discussed later), it puts pressure on the captive to properly price the unrelated business.

  A group captive allows smaller and/or less well capitalized entities to participate in a captive arrangement; it also provides for risk sharing, which may be a consideration for low frequency/high severity exposures. It does, however, require a sharing of information among the members, as well as a sharing of adverse experience. A group captive will often have a higher expense ratio than a single parent captive, as there can be more administration costs to run this type of program. It may, however, provide a better opportunity for acceleration of premium deductions than a single parent captive.

- Sponsored cell (or rent a captive) – In this arrangement, the sponsor, which may be a traditional insurer, establishes a captive and the participants use (or "rent") a cell in the captive (i.e., the sponsor's capital provides the financial backing for the business the cell owner wishes to place in the captive). Note that the sponsor is not an insured of the captive and the participants (or insureds) do not own or control the captive. These programs can be written on either a direct or reinsured basis. There are two general ways for participants to share in the captive results:

- Percentage participation – Each participant's share of captive overall profits/losses would be determined as a percentage of its premiums, its losses or some other pre-determined value. All participants' assets would be available for any losses that occur during the insured period.
- Protected (or segregated) cell captive – In a protected cell arrangement, the assets of a cell are protected from the liabilities of all other cells within the company; thus the participant's profit or loss is based on its own experience, subject to the level of "protection" provided by the reinsurance in place. Note that these "protected" vehicles are relatively new, so there is no real track record on how the protection would actually work or how it would hold up outside of the domicile.

A sponsored cell arrangement allows an entity to participate in a captive without any capital contribution. However, the expenses associated with this type of arrangement are typically higher, perhaps significantly, than for a single parent or group captive. The expense differential is driven by three factors:

1. The sponsor needs to be compensated for the use of its capital.
2. In the case of a protected cell captive, the cost of the "protection" would be passed on to the program participants.
3. Under some programs, the cell participants do not have the flexibility to select their own vendors (e.g., an investment manager) and may thus pay higher fees to the vendors selected by and/or owned by the sponsor.

- Risk retention group – A risk retention group (RRG) is a variant of a captive with a few key differences. RRGs were authorized under the Liability Risk Retention Act of 1986 to provide liability insurance (including products and medical malpractice, but specifically excluding workers compensation). All owners/insureds of an RRG must be engaged in businesses that have similar or related liability exposures. Both vertical and horizontal RRGs are permitted (vertical and horizontal are defined below). The RRG is an on-shore entity (and can not be domiciled off-shore). After being chartered in one state, an RRG can write insurance on-shore without having to become an admitted insurer in every state in which it does business. The RRG does have to register with any state other than the domiciliary state in which it plans to write insurance. This type of structure is useful for entities that cross state lines (e.g., for a healthcare system, which is providing insurance to physicians in a number of states). An RRG may cost more to operate than a captive, because it is often regulated more like a traditional insurance company than a captive insurance company and the capital requirements can be more significant. Unlike a captive, however, it does not require a front to write business domestically.

RRGs can either be owned directly by the entities insured by the RRG (horizontal RRG) or by a single entity that owns the RRG, which then has its members and owners as the insureds of the group (indirect or vertical ownership). RRGs must have at least two policy holders (unlike a single parent captive). The RRG can be organized in any form permitted by its domiciliary state, i.e., a stock company, a mutual company, or a reciprocal. A reciprocal RRG is an incorporated association, managed by an attorney-in-fact, which allows its members (or subscribers) to exchange contracts of "insurance". Depending on the reciprocal's structure (as determined by its bylaws), profits and losses can be allocated back to the subscribers' accounts.

The following table compares some key criteria with respect to the various types of captives.

Captive Type	Who Supplies Capital	Use of Front?	Off Shore?	Typical Users
Single Parent	Owners	Maybe	Maybe	Larger corporations, health care systems
Group Captive	Owners	Maybe	Maybe	Smaller corporations, universities
Sponsored Cell	Sponsor	Maybe	Maybe	Small corporations
Risk Retention Group	Owners	No	No	Health care systems, Affinity groups

There are two other vehicles, which are often referred to as captives, but which are not captives in the strict sense (i.e., the owners are not necessarily the insureds).

- Agency captive (or producer – owned reinsurance companies (PORCs)) – Insurers often offer captive participation to groups of agents to allow them to share in the underwriting profits produced by their accounts (in some cases the PORC is formed by agents and industry associations). In this scenario, the captive owners are not its insureds.
- Special purpose vehicles (SPVs) – These offshore vehicles can be used to securitize insurance exposures (e.g., cat bonds).

2.2 Domiciles

Captives are domiciled in venues that have passed legislation that allows captive formation, either on-shore (e.g., Vermont, South Carolina, Hawaii), or offshore (e.g., Cayman Islands (Cayman), Bermuda, Guernsey). Currently the most popular domiciles are Bermuda, Cayman, and Vermont; the 2002 Best's Captive Directory indicates that about two-thirds of the licensed captives are in five domiciles.

Domicile	Number of Captives	Net Written Premiums ($ in Billions)
Bermuda	1,625*	$28.8
Cayman Islands	665*	5.3
Vermont	674*	3.5
Guernsey	408	3.6
Luxemburg	280	2.7

Based on the 2002 Best's Captive Directory, updated with information from the domicile websites as available.
* Denotes updated data.

Domiciles vary with respect to their capital requirements, the level of regulatory oversight, and their accounting rules, among other items. These factors, as well as the captive owners' perception of the regulatory environment and the underlying captive infrastructure (e.g., the availability of captive managers, auditors, etc.) will influence the domicile choice. Other criteria considered in the decision on domicile include:

- The level of premium taxes
- The ability to discount liabilities
- Coverage lines permitted

In some instances, the choice of domicile is limited by the type of coverage the captive owner is considering; for example, writing employee benefits coverage under the United States Department of Labor Expedited Process Rules requires the use of a U.S. licensed captive.

It is important to note that a captive is considered an insurance company in its domicile; it is generally not considered an insurance company in the state (or states) where its parent/owner is located. Thus, in general, a captive can not perform any of the functions associated with the providing of insurance outside of its domicile (this limitation does not typically apply to an RRG). Depending on the structure of the insurance arrangement, however, some insurance company functions (such as claims) could be handled outside of the domicile (i.e., in the case of an indemnity policy whereby the captive agrees to indemnify the parent).

2.3 Reasons to Form a Captive

There are a number of reasons why organizations consider forming captives, including:

- Cost reduction - A key driver in captive formation is the potential to reduce insurance costs. Factors contributing to this include:
    - Better than average loss experience – An entity with good loss experience could have lower insurance costs using a captive because insurer pricing tools such as experience rating do not fully capture differentials in loss levels.
    - Lower expenses – A captive owner could negotiate more attractive terms than the expense level that is reflected in an insurance company's expense ratio as it may be able to purchase "unbundled" services from its vendors.
    - Retention of investment income – Investment income generated on unearned premiums and loss reserves is retained by the holder of these liabilities, which

would typically be the insurer under a traditional insurance program. A captive can retain the investment income on its unearned premium and reserves, potentially decreasing its parent's insurance costs.

- Improved cash flow – In addition to the retained investment income, a captive can provide its parent with improved cash flow through a more flexible premium payment plan or a more rapid payment of claims.

In spite of these potential advantages, a captive may not necessarily provide its parent with lower insurance costs, particularly if:

- Loss experience is worse than average
- Expenses are higher than an insurer's expense ratio
- Investment results are poor
- Premium is higher than competitive market pressures would indicate
- The capital invested in the captive could be better deployed within the corporation.

■ Accelerate tax deductions – An insurance company is allowed to deduct both paid losses and reserves (case plus IBNR), while a self-insurer is only allowed to deduct losses as they are paid. An organization may be able to use the insurance accounting treatment by establishing a captive (again this depends on the organizational structure) or participating in some form of a group captive. The determination of premium deductibility is complex, in part because insurance is not explicitly defined within the tax code. Factors considered in the evaluation of deductibility include:

— Risk assumption – Does the captive assume some risk, as supported by its business plan, or is the program a financing mechanism?

— Corporate structure – Are premiums paid to the captive by subsidiaries (e.g., a brother sister relationship), or the parent?

— Who owns the captive? Are the insureds different from the captive's owners?

- Increase capacity for its parent – A captive can provide coverage that may not be readily available from the commercial market (i.e., in that it is unique or in that it is not affordable). It can also provide coverage for gaps in the parent's program (i.e., a quota share percentage of an excess liability layer).
- Manage corporate retentions – Many entities use their captive for programs such as deductible buy downs; in these types of programs the parent may take on a fairly high deductible level (e.g., $250,000 or $500,000 per occurrence) but each division may be responsible for the first $25,000 or $50,000 per occurrence. The gap between the parental deductible and the division deductible can be funded in the captive, through either a pure risk sharing among divisions or based on some form of experience rating. The existence of the captive would allow the parent to have greater flexibility in its risk financing while dampening the impact of large claims on a single division's results.
- Centralize risk financing – A key benefit many organizations derive from a captive is that it provides a tool for the coordination of various risk financing arrangements and raises the profile of the risk management function to a higher level in the organization. A centralized focus can be a key element in controlling organizational risk financing.
- Direct access to reinsurance – A captive allows a company to access reinsurers directly, rather than relying solely on the brokered market. This has become a less compelling advantage since a number of reinsurers/excess insurers can be accessed directly and/or through the brokered market without the need for a captive. However, reinsurance is a critical component of a well-managed captive; in addition to per occurrence coverage, captives, particularly newly formed captives, will tend to purchase quota share and/or aggregate stop loss coverage, if available at an acceptable cost. These covers are helpful to newer captives to allow them to manage their exposure and build surplus.
- Support business partners or customers – Some organizations use their captives to provide "insurance" to their partners (e.g., vendors) or customers (e.g., warranty coverage). This can provide the owners with potential tax advantages (to the extent it is "third-party" business as defined by the Internal Revenue Service), risk diversification and underwriting profits, if the business is properly

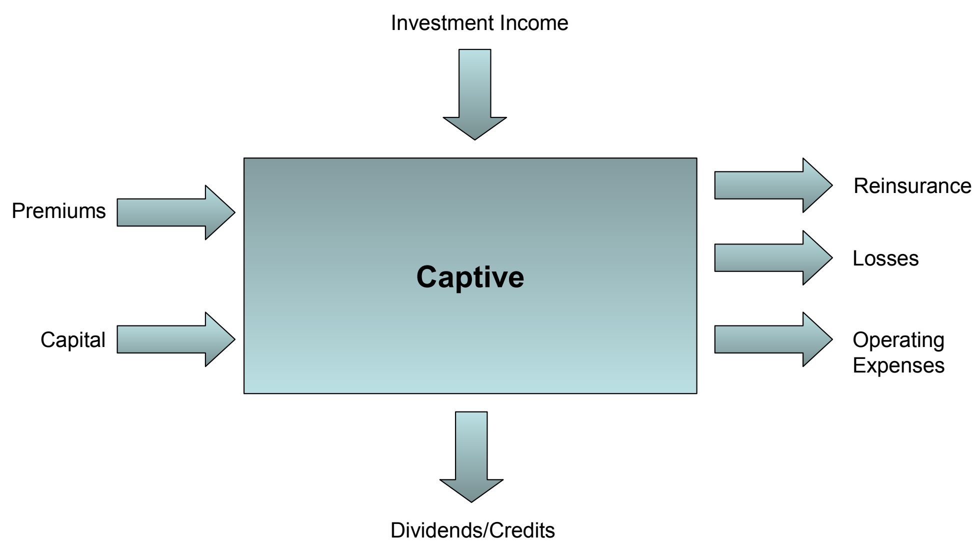
priced. It can also increase "loyalty" to the extent that the coverage offered is difficult or expensive to purchase in the commercial market.

3. RATEMAING ISSUES

An entity considering establishing a captive will typically evaluate the feasibility of this risk-financing vehicle. While a feasibility study considers a number of strategic issues in addition to losses and expenses (e.g., the impact of Federal income taxes, the viability of the business plan), this paper will focus on the loss and expense components of the analysis.

In its simplest terms, the premium charged by a captive should be sufficient to cover the expected losses and expenses associated with the coverage provided; appropriate pricing should consider the impact of expected investment income and a provision for adverse deviation, as discussed below.

The following chart shows simplified cash flows associated with a captive.



Pricing for a single parent captive tends to be based on a break-even profit assumption, whereas pricing for a group captive may have an assumed profit provision, which will be returned in policyholder dividends or premium credits should the experience be favorable.

Generally, more difficulties arise in developing captive premiums at the inception of the captive or when a new coverage is offered.

While the following analysis focuses on captive premiums, the approaches described can be used in the analysis of self-insured programs. The most significant differences between a funding analysis for an individual self-insured and a captive are:

- Loss focus – Typically the funding analysis for an individual self-insured will focus on the loss component, rather than the combination of losses and operating expenses. This is because the associated program expenses are often developed within the context of the risk manager's budgeting process and the amounts held as liabilities or in a trust (which could be used as a funding vehicle for a self-insured program) generally consider only losses and allocated loss adjustment expense (ALAE).
- Discounting – Many individual self-insureds do not consider discounting in the development of their self-insured accruals because it can be difficult, within a corporate structure, to develop a mechanism to credit back investment income. A risk manager

booking discounted liabilities increases the likelihood that future upward liability adjustments will be needed.

- Risk Margins – Often individual self-insureds establish liabilities at an expected level; i.e., they do not include a provision for adverse deviation. In some cases this approach is driven by auditor considerations (e.g., the FAS 5 criteria, related to accruing loss contingencies, of probable and estimable). In other cases, the decision to book at an expected level is cost driven.

For a workers compensation self-insured group (SIG) the ratemaking analysis will typically involve establishing a deviation (from published premiums or loss costs) for the SIG; this can involve a consideration of both the loss and operating expense components.

3.1 Data Issues

There is a wide variety in the quality of the loss and exposure data available for captive pricing analyses. Some entities will have complete historical loss data spanning a number of years; in other cases the data will consist of loss runs by participant (e.g., as can be the case when evaluating the experience for a group of potential captive owners who previously had individual insured programs). Some common problems with an initial data submission include:

- Exposures without losses – The loss information provided may not include historical information for all exposures (e.g., for a group with 30 potential members, loss runs may be provided for only 20 of the members). There are two ways to address this issue; obtain the missing data or develop the analysis using only the exposures that provided loss data. While the latter approach may reduce the credibility of the results, the assumption that the missing data reflects no losses is likely optimistic. Note, also, that this approach assumes that the loss experience for potential members with no loss data is similar to that of potential members with loss data. In some situations, there may be additional data that could be used to check the validity of this assumption (e.g., workers compensation experience modification factors).

- No closed claim data - Another source of incompleteness relates to data that only includes open claims; often loss runs, particularly from expired programs, will show activity on pending claims and/or claims closed within the evaluation period. Open claim data can be used to project ultimate losses; however it will typically produce a more volatile result than more standard development methods.

- Combined coverage data – Entities that have coverage on a combined lines basis will often submit loss information that does not separate the claims by coverage. In order to develop useable loss projections the loss and exposure data needs to be split by coverage. To the extent the data cannot be split out, the actuary can use industry statistics to split the data into its component pieces (e.g., a common assumption is that general liability would represent between 5 to 10% of a combined medical/general liability program). This approach may not accurately measure the exposure of a specific entity.

- Incomplete exposure data – Often exposure data is provided for the latest year or two, while the loss data may cover a longer period. Typically, the captive organizer may be able to provide growth assumptions that can be used to develop estimates of historical exposures. For certain companies, this approach will produce a reasonable estimate of historical exposures. In cases where there have been significant changes (for example, through an acquisition or divestiture), this approach may create a mis-match between exposures and losses. Depending on the volume of losses, the limitations in the exposure data can have a significant impact on the credibility of the analysis. A typical experience period is five to seven years; for a larger volume of losses in a short tail line, three complete years of data could be sufficient. Conversely, for a high severity low frequency exposure, a ten year experience period may not be fully credible.

- Inconsistent exposure data – Inconsistency in exposure data is often found when a group program is being analyzed (i.e., a group of nursing homes considering captive formation), though it can also be found in a single parent situation. The differences may be relatively simple, such as what is being counted (e.g., what is a "bed") and is the term defined in the same manner across the group (e.g., skilled care, assisted living and independent living beds may all be considered to be long term care beds, but they represent significantly different exposures. In other cases, the differences may involve more complex issues, such as what is included in the data (for example, payroll can be straight payroll, payroll including overtime, payroll with benefits, etc.). Sometimes the exposure data provided is not internally consistent (e.g., consistently increasing premiums with decreasing payroll). While there are some scenarios where this may be a plausible relationship (e.g., where payroll reductions result in deteriorating loss experience, with a parallel increase in workers compensation experience mods), often this reflects incompleteness or inconsistency in the exposure data.

- Claim count data – Claim count data is often not provided or is not available at a useful level, which limits the types of analysis that can be performed. Also, if claim count data is provided from a number of claims handlers, the compiled claim count data may have limited use because different insurance companies and different third party administrators (TPAs) often have different criteria for the establishment of claim files. This data issue becomes more problematic when there are multiple insurance companies and TPAs and/or changes in claims handling practices.

- Partial loss data – Occasionally an entity will provide loss data that may not consistently include ALAE. This is often true for coverages such as employment practices, where the ALAE may be tracked outside of the risk management function (i.e., in legal). To provide a valid comparison to an insurance product, it is necessary to reflect the ALAE data in the analysis since it is generally a significant portion of ultimate costs for this type of exposure.

3.2 Industry Statistics

Some of the issues described above can be resolved through the back and forth dialog of the data collection process; other issues may prove more difficult. For this reason, many captive pricing analyses typically rely more on industry statistics than would the pricing for an insurer's products.

The more commonly used industry statistics include:

- Benchmark loss development patterns – Many entities do not retain the historical data necessary to construct development patterns and/or they are not able to obtain this information from carriers with expired programs. In other cases, the data triangulation may lack credibility (e.g., for a low frequency/high severity exposure or where there have been multiple carriers insuring the risk). Typical sources for benchmark patterns include compiled industry data (e.g., Best's Aggregates & Averages) or publicly available rate filings. As with any benchmark, it is important to consider how the entity's data fits the benchmark statistics used. For some coverages, publicly available data is limited and/or non-existent (e.g., umbrella liability) so the benchmark pattern may be highly judgmental.

- Industry size of loss curves – A larger entity may not have fully credible data beyond a certain limit (e.g., $500,000 per occurrence) while a smaller entity's data may not be credible beyond significantly lower limits (e.g., $50,000 per occurrence). While this issue is more critical in low frequency/high severity lines (such as professional liability), it often is a factor in pricing captive exposures in more typical lines (e.g., auto liability). In many captive pricing analyses, losses are limited (for example, to $100,000 per occurrence) and the exposure above this limit is estimated using industry increased limits factors that can be obtained from rate filings.

- Trend Factors – Most entities do not have sufficient data that would allow for the determination of credible trend factors. Again, this is more of an issue for low frequency/high severity lines. For entities with a reasonable volume of stable data, their own trend information can supplement industry data; for other entities, industry trend data would be used directly. Again, the typical source of trend data is industry rate filings. These can be supplemented with Consumer Price Index (CPI) data, the Masterson Index (as published by A.M. Best), and other economic trend data.

- Industry loss costs – An entity's historical loss experience may be too sparse or volatile to provide a reliable indication of the potential exposure to loss. Industry loss costs can be used as a supplement to entity data; these industry loss costs are often developed from rate filings. Note that considerable judgment can be necessary when using industry loss costs particularly in lines where there has been a significant change in the market.

- Statutory changes – For certain coverages (e.g., workers compensation or professional liability) statutory changes can significantly impact future costs. For workers compensation, in the absence of a law reform, these annual changes tend to be 1% or less, driven by medical and/or wage inflation. For professional liability, these can be significant, depending on the type of tort reforms enacted. For captive analyses, the actuary will typically need to rely on published evaluations of law changes, with judgmental adjustments to reflect differences in an employer's workers compensation program (e.g., a different level of use of utilization review) or to adjust for known limitations in the data used to price a tort reform package (e.g., the applicability of the state data used in the analysis to the state in question, given the wide variation in the tort environment by state).

4. RATEMAKING EXAMPLES

The following section includes three examples of potential ratemaking approaches that could be used in a captive scenario. Note that in each of the examples (Exhibits 1 through 3), the analysis is developed from the last sheet forward to Sheet 1, which summarizes the results.

4.1 Example One

A single parent captive is considering writing an indemnification policy for its self-insured workers compensation program where the self-insurer retains the first $500,000 of any occurrence. The company has an existing captive and adding this coverage would allow more diversification in the captive. The company has a relatively high volume of claims, and the largest claim reported to date is valued at $600,000. Five years of loss and exposure information was available (See Sheet 7 of Exhibit 1, which summarizes the data).

The actual analysis is relatively straightforward. First ultimate losses limited to $100,000 per occurrence are developed using the company's data and industry loss development patterns. Two projection methods are used (incurred and paid loss development) and ultimate losses limited to $100,000 per occurrence are selected for each year (see Sheet 3 of Exhibit 1). The estimated ultimate losses are trended and adjusted for benefit level changes and compared to payroll to calculate a limited pure premium for each year (see Sheet 2 of Exhibit 1). A limited pure premium is selected based on the historical results. In this exhibit, frequencies and severities are also calculated to check the reasonability of the projections. Expected losses limited to $100,000 per occurrence for the upcoming year are calculated in Exhibit 1, Sheet 1 by multiplying the selected limited pure premium by the projected exposures (payroll).

The indication is then adjusted to reflect the program retention ($500,000 per occurrence) using an increased limits factor (ILF). Given the limitations in the data (e.g., complete individual claim detail was not available since the large loss summary only captured information on claims valued at $100,000 or more), the company's experience cannot be used directly to develop ILFs. In Sheet 6 of Exhibit 1, we compare the company's implied ILF to industry data. The calculations shown in Exhibit 1, Sheet 6 provide only general guidance as to the entity's large loss experience (in this case suggesting that the company's large loss experience is somewhat more favorable than is implied by industry data). Note that this approach gives a very "macro" sense of how the company's large loss experience compares to industry large loss experience. A more traditional approach involves developing and trending individual claim data to calculate ILFs; this methodology requires a large volume of claims, preferably including all claims rather than just those over a certain dollar limit. Given this data constraint, captive analyses tend to rely more on published ILFs. The selected ILF used in the projection on Exhibit 1, Sheet 1 relies mainly on industry data, given the limited credibility of the company's historical large loss experience.

We then calculate a risk margin and adjust the indication to reflect discounting and operating expenses in Exhibit 1, Sheet 1. The parent has determined that it will use a 75% confidence level factor in its captive pricing, as this is acceptable to the domicile's regulators.

The company and industry data are used to develop the frequency and severity parameters for a loss simulation in Exhibit 1, Sheet 2. To develop the frequency estimates, reported and closed claim counts are projected to an ultimate basis using benchmark patterns (see Exhibit 1, Sheet 4). Given the differences in company and TPA approaches to opening claim files, there is little industry data to use to develop claim count patterns, which means that the actuary may need to develop patterns from the experience of similar entities. In deriving the simulation parameters, medical only claims are excluded given their low average severity (less than $500 per claim typically). This provides a "truer" picture of potential variation, since the inclusion of a large number of low severity claims would dampen the results (i.e., produce a narrower risk margin). For liability lines, excluding closed no payment claims would have a similar impact. Exhibit 1, Sheet 2 shows the details of the adjustment.

For modeling purposes, a Poisson distribution was assumed for frequency and a lognormal distribution was assumed for severity; the coefficient of variation (one of the inputs for the lognormal distribution) was estimated based on industry size of loss distributions. Note also that the severity derived in this calculation is limited to $100,000; for simulation purposes, this often needs to be adjusted to an unlimited basis. A common problem in deriving simulation parameters for feasibility studies relates to calculating an unlimited severity. Even if the data is provided on an "unlimited" basis, there is generally not a sufficient volume of large claims for the data to be actually unlimited. It is important to determine the "implied" limit in the data and then adjust the calculated severity to an unlimited basis (i.e., through the application of ILFs); otherwise the calculated risk margins may be understated.

Present value is reflected using an industry loss payment pattern and an assumed investment yield (see Exhibit 1, Sheet 5). Captive pricing often considers the time value of money, in part because many of the major captive domiciles allow discounting of reserves and/or prospective funding. In determining an appropriate discount rate, the actuary typically relies on input from the parent, and/or the captive's investment advisor. In some domiciles (e.g., Caymans) captives are allowed to have a greater percentage of equity investments than U.S. statutory rules would allow, which could have a favorable impact on the assumed investment yield.

Operating expenses are then added to the discounted 75% confidence level losses to determine the captive premium shown on Exhibit 1, Sheet 1. Typical expenses could include:

- Captive management (producing the captive financials, dealing with regulators, financial reporting/MIS
- Excess insurance or reinsurance, potentially including some form of aggregate coverage.
- Claims handling, if not included in management
- Fronting fees
- Audit
- Actuarial
- Legal
- Taxes – These include state premium taxes or possibly direct placement/self-procurement taxes, federal excise tax (if "insurance premiums" are paid to an offshore domicile), federal income taxes.
- Investment expenses (if not netted out of investment income)
- Travel costs (for Board of Directors meetings)

- Letter of credit (LOC) costs if needed; generally a front will require collateral and LOCs are often used to meet the collateral requirements. LOCs are often used as part of a captive's capitalization.
- Risk management/loss control services (if any)
- Other expenses: brokerage commissions/fees, any sponsorship/endorsement fee (probably not for single parent), domicile charges (amount and basis varies by domicile), bond fees, D&O insurance (for the Board of Directors), etc.

Annual expenses for a single parent captive typically range from $50,000 to $150,000 (excluding excess insurance or reinsurance). Note that this range contemplates a fairly generic program; complex multi-line captives will likely have higher operating expenses.

The expenses included in Exhibit 1, Sheet 1 are:

- Excess insurance; and
- A pro-rata amount of total program expenses (e.g., captive management, audit). In this example the policy is assigned a 10% pro-rata share of expenses; the remainder of the total program expenses is allocated to the other coverages written in the captive. The 10% allocation was determined by comparing this coverage's expected losses to the total captive expected losses. A more refined allocation procedure could be used (i.e., to reflect differences in the various expense components by coverage). Exhibit 1, Sheet 8 shows the details of the expense calculation.

Many captive programs incorporate a retrospectively rated premium (retro-rated) feature, where the premium reflects the insureds' loss experience subject to minimums and maximums. This approach provides an advantage to the captive in the event that loss experience is adverse; however, it may have a negative impact on the acceleration of tax deductibility.

At future evaluation dates, the above analysis could be modified in a number of ways to better reflect the entity's loss experience, including:

- Company specific loss development triangles could be created. Note that this would be a longer term initiative (unless historical loss valuations were available). It may also be difficult to capture future valuations of run-off programs (in this case, pre-captive experience) unless there is a mechanism for the company to obtain this data from its prior carriers.

Developing a company specific loss payment pattern would also effect the discounting calculation and would potentially imply a lower discount amount (i.e., if captive losses are paid more quickly than implied by industry benchmarks)).

- A more robust adjustment to industry size of loss curves could be developed, which would affect both the increased limits adjustment and the calculation of risk margins. It could also impact the captive's retention level for this coverage, as it would allow a better comparison of the cost of retaining the exposure relative to reinsuring it.
- The lost time/medical only split could be modified to reflect emerging experience.

4.2 Example Two

The next scenario involves a group of four physician practices seeking to form a captive to cover their professional liability exposures. There are two factors driving their interest in a captive – their loss experience has been extremely favorable and their premium expenses have increased significantly. Market conditions would suggest that the lowest attachment for excess insurance/reinsurance for the proposed captive is likely to be between $1 million and $2 million per occurrence. The physicians are presently in a first dollar program written on an occurrence form.

Given the level of exposure (i.e., the retention level of at least $1 million per occurrence) under consideration, a key question to consider is the credibility of the data; is it reasonable to assume that the loss experience is fully credible or should it be supplemented with other sources of data? The volume of large claims in the data shown in Exhibit 2, Sheet 4 (a large loss listing) does not provide a true picture of ultimate losses at the levels of coverage being considered for the captive, given the volume and level of large claim activity (for example, there are only two claims in excess of $500,000). There are also some limitations in the overall data provided, which are summarized by accident year on Exhibit 2, Sheet 5 (this exhibit also includes diagnostic statistics calculated to provide insight on data "reasonability"). Some observations include the following:

- Exposure information is not provided for all policy years

- The trends in loss information are not necessarily consistent with the trends in exposure data. (Note the loss level in the 2000 year relative to the prior and subsequent years; it is over 25% greater, while the exposure level is relatively consistent). Given the nature of the exposure (professional liability) this fact on its own may not be indicative of significant data issues.
- Average values of open claims do not track average paids, nor does frequency track loss volume
- The data quality appears to vary by entity.

In reviewing this submission the actuary would need to try to resolve these questions and/or obtain additional data. To the extent these issues cannot be resolved, an approach would be to exclude questionable data and develop the analysis based on a smaller volume of apparently more reliable data. However, the corresponding reduction in data credibility may limit the appeal of this approach.

Exhibit 2, Sheet 3 details the pricing approach used, which is based on an "experience rating" model. The first step is to evaluate the data to determine at what loss limit it is credible (in the example, we have assumed losses limited to $100,000 per occurrence are fully credible). Then estimated ultimate losses for each accident year are calculated by multiplying the basic limit incurred losses by loss development factors. These ultimate losses are divided by exposures on a base class basis; in the example, we have assumed that all of the physicians practice in the same specialty. Often, it would be necessary to adjust the exposures to a common (or base class) basis by multiplying the number of physicians for each specialty by classification factors, which reflect the relative "riskiness" of each specialty. These class factors can be obtained from rating manuals; note that there is significant variation in the class rating schemes using by different carriers. This calculation produces a developed loss cost per base class physician, which is then compared to an industry expected loss cost. The industry expected loss cost could be derived from rate filings, or the experience of similar exposures. The loss development factors used in this Exhibit are also based on industry experience.

The actual loss costs are compared with the expected loss costs to determine an experience modification factor (experience mod) for each accident year. This calculation could also be done on a paid basis, but given the length of the expected payment pattern for professional liability, the paid comparison may be less meaningful than the reported loss comparison presented herein. The individual accident year results are then weighted (using exposures and reporting patterns as a proxy for the implied credibility of each year) to calculate an overall weighted average experience mod factor.

The credibility of the loss experience is determined by calculating a credibility factor; in this case, it is based on the number of insured physicians, on a base class basis. A full credibility standard of 40,000 is used in the example. A credibility weighted experience mod is calculated (reflecting a unity factor for the balance of credibility), and based on this calculation, an experience mod is selected. The selected experience mod is applied to the industry expected loss cost to calculate an experience-modified loss cost. The product of the experience-modified loss cost (from Exhibit 2, Sheet 3) and projected exposures (on a base class basis) is estimated losses for the forecast period (see Exhibit 2, Sheet 2).

To this point, the analysis has been performed on an accident year (or occurrence) basis. Given that reinsurance for this coverage is generally written on a claims-made basis, the coverage through the captive will be provided on a claims-made basis (using the same coverage form reduces coverage gaps that can arise when there are changes in retention prospectively). Since the physicians have historically been insured on an occurrence basis, they do not need to purchase coverage for prior exposures (e.g., tail coverage from their current insurer or nose coverage from the captive). As such, the initial captive premium would reflect first year claims-made coverage (that is, the captive would cover all claims reported in the first policy period occurring on or after the retro date (in this case, it would be the effective date of the policy)). Accounting guidelines may also suggest that a tail premium be included to cover claims reported subsequent to the expiration of the policy period. The accident year losses are converted to a claims-made basis on Exhibit 2, Sheet 2 through the application of a claims-made factor (again, based on industry data).

An increased limits factor is then used to adjust the losses to the appropriate retention level. In this example, where three retentions are evaluated, the ILFs are based solely on industry data due to the credibility (or lack thereof) of the physician groups' large loss volume. Similarly to Example One, risk margins and discounting are incorporated and expenses are added to the loss projection to estimate premiums (see Exhibit 2, Sheet 2). However, there are three important differences from Example One:

- The risk margin parameters are developed entirely from industry data, given the volume of the historic data.
- The entire program expenses are reflected in the premium, since the proposed captive contemplates a single coverage (Exhibit 2, Sheet 7 details the expense components, excluding profit).
- The expenses include a profit loading, which could be returned to the members in the form of a dividend, should the experience be favorable.

Exhibit 2, Sheet 1 shows the allocation of premium by physician group. Column (4) of this exhibit shows an allocation based on exposures, while Columns (6) and (8) show allocations based on counts and incurred losses, respectively. The final allocation is based on equal weightings of the three percentages. Note that this weighting is judgmental; other weights could be used. In developing the allocation methodology, factors to consider include:

- The level of risk sharing among group members; a loss-sensitive allocation system generally implies less risk sharing among the members.
- The impact of loss control and risk management – to the extent that these two factors can influence experience, a loss-sensitive allocation can have a long-term favorable impact. If losses are more fortuitous, a loss-sensitive allocation may be considered punitive.
- The variation in member size – if a group is comprised of large and small members, an exposure-based allocation may not reflect economies of scale that could be attributed to a larger member. In practice, groups where members vary significantly in size (particularly where there is one large member and a number of small members) may find it difficult to develop a "fair" allocation.

After the premium allocations are developed, the individual premium is increased by a factor of 50% to incorporate an initial capital contribution. The combination of the risk margin and the initial capital contribution are estimated to reflect a 90% confidence level. Given the level of retention under consideration and the line of coverage (professional liability), a start-up captive would typically fund at this confidence level.

Over time, captive experience could be incorporated into the analysis, but at a much slower rate than in Example One. This difference arises due to the nature of the coverages considered in the examples (workers compensation in Example One at $500,000 per occurrence limits vs. professional liability in Example Two at $1,000,000 per occurrence limits). For a number of years, it would be necessary to rely heavily on industry development statistics and size of loss curves, as well as industry loss costs.

4.3 Example Three

In this situation a company is considering writing coverage in its captive for a new exposure that is not underwritten and/or reasonably priced in the insurance market. Because of the novelty of the exposure there is little or no industry loss data available, which means that pricing would need to rely on non-insurance data.

The first step in evaluating this type of exposure is understanding the process by which an insured event would generate a claim and then modeling the process. This could involve interviewing the potential insured and obtaining external data. An example of this situation would be a healthcare entity that is considering offering coverage through its captive for workers compensation claims that could arise from immunizing healthcare workers for smallpox.

For purposes of this example we look at frequency and severity separately; in some types of these analyses it would be necessary to develop projections on a combined basis, due to limitations in available data. To simplify the modeling process, we also assume that there would only be two ways in which a claim could arise:

- A vaccinated worker contracted smallpox; or
- A vaccinated worker infected a co-worker.

To develop a claim frequency projection, it would be necessary to compile exposure data for the potential insured. This would be combined with industry frequency data (in this example, the Center for Disease Control (CDC) website could provide a range of useful input) to estimate a claim projection for the insured's program. We estimate two claim frequencies on Exhibit 3, Sheet 3 (the claim frequency for a direct exposure (a worker contracting smallpox directly from the vaccination) and the claim frequency for an indirect exposure (a worker contracting smallpox from a fellow worker having been vaccinated)) and combine the implied ultimate claims from the two potential exposure sources (i.e., we are assuming the severity of a claim will not vary regardless of how the claimant was exposed). The key variables underlying the claim frequency projection are:

- Projected exposure. In this example payroll was converted to number of employees, since the publicly available data related frequency to an employee headcount. The conversion assumed an average salary per employee. Note that this simplified example does not consider exposure differences among categories of employees (i.e., physicians vs. administrative staff); to incorporate such a differential, it would also be necessary to consider salary differentials among the employee groups when converting payroll to headcount.
- The percentage of workers vaccinated and the estimated percentage of non-vaccinated workers exposed to vaccinated workers; this information was provided by the healthcare system.
- The estimated percentage of vaccinated workers contracting smallpox and the estimated percentage of non-vaccinated workers contracting smallpox; this was based both on industry information and input from the healthcare system.
- An "interaction effect". This factor is essentially a modifier which is used to adjust the projection to reflect an actual or perceived difference in the potential insured's exposure relative to that implied from the publicly available data (i.e., if it was believed that the level of interaction among employees could result in higher or lower infection rates than external data would suggest). In Exhibit 3, Sheet 3 it was assumed that the interaction effect would increase the number of claims by 20%.

The estimated claim counts for the direct exposure are calculated on Exhibit 3, Sheet 3 as the product of the estimated headcount, the percentage of workers vaccinated, and the percentage of vaccinated workers contracting smallpox. Similarly, the estimated claim counts for the indirect exposure is calculated as the product of the estimated headcount, the complement of the percentage of workers vaccinated (to determine non-vaccinated workers), the estimated percentage of non-vaccinated workers exposed to vaccinated workers, the estimated percentage of non-vaccinated workers contracting smallpox and the interaction effect. The total projected claims on Exhibit 3, Sheet 3 are the sum of the projected claims for the two exposures.

Each worker infected with smallpox will experience varying levels of disease and associated costs. To simplify the example, we assume one of three outcomes (using a workers compensation industry claim categorization).

- Outcome A - A fatal claim (the claimant dies within two weeks of exposure);
- Outcome B - A permanent total claim (the claimant is permanently unable to work); and
- Outcome C - A temporary total claim (the claimant is out of work for eight weeks). Given the nature of the disease, we assume that there will be no minor claims (e.g., medical only claims).

Percentage probabilities are assigned to each outcome based on external data and input from the healthcare system and estimated severities are developed for each of the scenarios. A claim severity for each outcome is shown in Exhibit 3, Sheet 2; this severity reflects the estimated indemnity and medical (both current and future) costs of each outcome. Key inputs include the assumed wage level, associated medical costs, future wage and medical trends, and the potential for benefits for dependents. The analysis of severity could be further refined to reflect a wider range of potential outcomes.

An overall estimated severity is determined by calculating the weighted average of the estimated cost of the three outcomes. The frequency and severity assumptions are then combined to calculate expected losses in Exhibit 3, Sheet 1. As in the prior examples, the expected losses are adjusted to reflect discounting, risk margins and operating expenses. Given the nature of the coverage, the fact that it is an additional coverage for the captive (so that the additional expenses are more of a frictional cost) and that no excess insurance or reinsurance will be purchased, the associated expense level is relatively minimal as compared to Example One or Two. Note that the payment pattern used in the discount calculation (Exhibit 3, Sheet 4) reflects the projected cash flows associated with each outcome, rather than an aggregate industry payment pattern.

In the absence of an actual incident, it may not be possible to further refine this analysis in subsequent years; thus, the potential captive premiums are primarily a function of the underlying assumptions and external data.

5. FINANCIAL CONSIDERATIONS OF A CAPTIVE

Developing captive premiums that reflect potential profit provisions and/or incorporate risk margins or capital contributions increase the financial strength of the captive and offer a number of advantages over a long-term horizon. These include:

- Enhancing the flexibility to change the program retention in response to market conditions
- Increasing the ability to raise premiums (i.e., by adding new members to a group captive or adding additional coverage to a single parent captive)
- Providing the flexibility to support a higher than average level of claim payments in a single year without liquidating assets
- Positioning the captive to meet solvency requirements of the domicile or a rating agency (as needed).

For a captive, the premium analysis needs to be considered in the context of the captive's financial position. Some of the key financial ratios to consider in evaluating the financial strength of a captive include:

- The premium to surplus ratio – This leverage ratio reflects a company's exposure to pricing errors; for example, if a company's premium to surplus ratio were 2:1, a 10% underestimate on premiums would have a 20% effect on surplus. This is probably the most commonly used leverage ratio, though there is not necessarily one "right" ratio. Factors to consider in evaluating a captive's premium leverage ratio include the type of business (exposure, policy form and limits offered), the relative adequacy of its pricing, and its approach to loss reserving. According to Tillinghast's Recognized and Accepted Captive Standards (TRACS), captives in general tend to have lower premium to surplus ratios than commercial insurers, because they often have higher risk retention to surplus ratios. A range of "normal" leverage ratios for captives is from 1:1 to 5:1, although it can be greater in some offshore domiciles.
- The reserves to surplus ratio – The reserve leverage ratio measures a company's exposure to reserving errors and again there is not necessarily a "right" ratio for a captive. In evaluating this ratio, it is important to consider the level of reinsurance used

by the captive and its approach to establishing reserves. Similar considerations apply to this ratio as to the premium to surplus ratio. A range of reserve to surplus ratios for captives is 3:1 to 5:1. At higher leverage ratios, a relatively small increase in reserve levels would have a significant impact on surplus.

Although large premium and reserve leverage ratios can be negative from the financial perspective of captive, it does not necessarily follow that low leverage ratios are positive. A low leverage ratio could imply that the captive is adequately priced or reserved; it could also imply that the captive is overcapitalized and that the "excess" capital could be put to better use within the organization. Conversely high leverage ratios could indicate that the captive's pricing and/or reserving is stronger (more conservative) than average.

- Risk retention to surplus ratio – A number of domiciles use the "10% rule" (i.e., a company may not expose more than 10% of its surplus to any single risk or loss). This ratio is considerably higher than the risk retention level of a large insurer. Many captive owners believe that captives should be risk takers, with the understanding that they may need to contribute additional premium or surplus if experience is adverse. Risk retention ratios of captives can range from 10% to over 200%, depending on the coverage, membership structure (i.e., single versus group) and domicile.

Any analysis of a captive's financial position also needs to consider its use of reinsurance. Typical ways reinsurance is used in a captive program include:

- Protection from catastrophe losses (either per occurrence or in aggregate) – purchasing reinsurance gives the captive more stability with respect to income fluctuations and/or solvency.

- Providing capacity – Captives often provide high limits of coverage (relative to the captive's surplus), particularly in low frequency/high severity lines. Purchasing reinsurance allows a captive to provide these limits while protecting the captive's surplus position.

- Supporting growth – A group captive or RRG may purchase reinsurance to facilitate growth because in the short run growth will have a negative impact on the captive's leverage ratios. The impact is assumed to be short run because it is assumed that the "business growth" is adequately priced. If this assumption does not hold, the purchase of reinsurance is more critical (assuming the reinsurer does not question this use of its surplus), or more importantly, the growth strategy should be re-evaluated.

- Providing an exit strategy – A captive withdrawing from some or all of the coverages offered will often transfer the remaining liabilities to a reinsurer through a LPT.

There are a number of other factors to consider in evaluating a captive's financial position, such as its investment portfolio, but they are beyond the scope of this paper.

6. CONCLUSION

Captives and alternative vehicles will likely continue to represent a large component of the risk financing market. It is likely that captive owners will continue to use these vehicles to finance new and different exposures, in addition to the more traditional coverage lines. This growth represents both a challenge and an opportunity to actuaries.

7. REFERENCES

[1] John D. Long & Davis W. Gregg, *"Property and Liability Insurance Handbook"*, 1965.

[2] A.M. Best, "*Bests Captive Directory"*, 2002.

[3] IRMI, *"Captive Insurance Company Reports"*, 2002-2004

[4] Towers Perrin *"Tillinghast Recognized and Accepted Captive Standard* (*TRACS): The Road to Successful Captive Management"*, 2004

[5] Towers Perrin, "*Captives 101", Managing Cost and Risk,* 2004

[6] Towers Perrin, *"Captive Insurance Company Glossary"*, 2004

Biography of the Author

Ann M. Conway, FCAS, MAAA, is a Consulting Actuary with the Tillinghast business of Towers Perrin in the firm's Boston office. She is a principal of Towers Perrin. She is a graduate of the Massachusetts Institute of Technology. She is a Fellow of the Casualty Actuarial Society (CAS) and a member of the American Academy of Actuaries (AAA). Ms. Conway is a member of various CAS Committees, including the Nominating Committee, the Long Range Planning Committee and the Regional Affiliates Committee. She previously published a paper in the Ratemaking Call Paper Program: "An Approach to Ratemaking for Self-Insured Workers' Compensation for Both Individuals and Groups" (1994).

Example One Exhibit 1
Projection of 2005 Premium Sheet 1

1. Estimated Payroll (00s)	$1,681,000	
2. Selected 2005 Pure Premium	2.90	
3. Increased Limits Factor	1.510	
4. Expected 2005 Ultimate Losses (000's)	$7,361	
5. Discount Factor @ 5%	0.8628	
6. Discounted Expected 2005 Ultimate Losses (000's)	$6,351	
7. Risk Margin at		
a) 75% Confidence Level	1.10	
b) 90% Confidence Level	1.30	
c) 95% Confidence Level	1.50	
8. Estimated Expenses (000's)	$225	
9. Estimated Premium (000's) at	Nominal	Discounted
a) Expected Level	$7,586	$6,576
b) 75% Confidence Level	8,322	7,211
c) 90% Confidence Level	9,795	8,482
d) 95% Confidence Level	11,267	9,752

Notes:

(1) Assumes 5% annual growth from 2003 level.
(2) From Exhibit 1, Sheet 2.
(3) From Exhibit 1, Sheet 6.
(4) (1) x (2) x (3) / 1000.
(5) From Exhibit 1, Sheet 5.
(6) (4) x (5).
(7) Based on simulation of Company experience.
(8) Provided by Company. See Exhibit 1, Sheet 8.
(9) (4) or (6) (for discounted) x (7) (for higher confidence levels) + (8).

Example One
Projection of 2005 Pure Premium - Limited to $100K

Exhibit 1
Sheet 2

Accident Year	Estimated Ultimate Loss (000s)	Trend/ Benefit Factor	Trended Ultimate Loss (000s)	Payroll (00s)	Estimated Pure Premium	Estimated Ultimate Counts	Estimated Frequency	Estimated Severity
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1999	$2,790	1.268	$3,536	$1,400,000	$2.53	610	0.436	$5,797
2000	2,880	1.218	3,509	1,425,000	2.46	630	0.442	5,570
2001	3,560	1.171	4,170	1,480,000	2.82	715	0.483	5,831
2002	3,980	1.126	4,481	1,500,000	2.99	760	0.507	5,896
2003	3,830	1.082	4,145	1,525,000	2.72	800	0.525	5,181
Total	$17,040		$19,841	$7,330,000	$2.71	3,515	0.480	$5,645
		(10) Selected			$2.90		0.505	$5,800
		(11) Adjusted to eliminate med-only claims					0.126	$17,400

Notes:

(2)	From Exhibit 1, Sheet 3.
(3)	Based on industry data.
(4)	(2) x (3).
(5)	From Exhibit 1, Sheet 7.
(6)	(4) / (5).
(7)	From Exhibit 1, Sheet 4.
(8)	(7) x 1000 / (5).
(9)	(4) x 1000 / (7).
(10)	Selected judgmentally.
(11)	(10), adjusted to reflect an assumed med-only percentage of 75% of claims and 25% of losses.

Example One
Projection of Ultimate Losses Limited to $100K (000s)

Exhibit 1
Sheet 3

Accident Year	Losses (000's)	LDF to Ultimate	Estimated Ultimate Losses	Losses (000's)	LDF to Ultimate	Estimated Ultimate Losses	Selected Ultimate Losses
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1999	$2,650	1.071	$2,837	$2,200	1.250	$2,750	$2,790
2000	2,650	1.092	2,894	2,125	1.351	2,872	2,880
2001	3,000	1.158	3,473	2,350	1.550	3,643	3,560
2002	3,300	1.273	4,202	1,800	2.083	3,750	3,980
2003	2,525	1.528	3,858	800	4.762	3,810	3,830
Total	$14,125		$17,263	$9,275		$16,825	$17,040

Notes:

(2),(5)	From Exhibit 1, Sheet 7.
(3),(6)	Based on Company and industry data.
(4)	(2) x (3).
(7)	(5) x (6).
(8)	Selected judgmentally.

Example One
Projection of Ultimate Counts

Exhibit 1
Sheet 4

Accident Year	Reported Counts	LDF to Ultimate	Estimated Ultimate Counts	Closed Counts	LDF to Ultimate	Estimated Ultimate Counts	Selected Ultimate Counts
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1999	600	1.001	601	575	1.075	618	610
2000	625	1.005	628	550	1.150	633	630
2001	700	1.010	707	600	1.200	720	715
2002	725	1.030	747	550	1.400	770	760
2003	725	1.100	798	400	2.000	800	800
Total	3,375		3,480	2,675		3,541	3,515

Notes:

(2),(5)	Provided by Company.
(3),(6)	Based on Company and industry data.
(4)	(2) x (3).
(7)	(5) x (6).
(8)	Selected judgmentally.

Example One
Calculation of Discount Factor
Assumed Discount Rate of 5%

Exhibit 1
Sheet 5

Year (t)	Percent Paid in Year (t)	Percent Unpaid at end of Year (t)	Present Value of Payments in Year (t)	Avg. Disc. Factor for Remaining Payments at end of Year (t)
(1)	(2)	(3)	(4)	(5)
0		100.00		0.8628
1	21.00	79.00	20.4939	0.8744
2	27.00	52.00	25.0946	0.8628
3	16.50	35.50	14.6053	0.8507
4	9.50	26.00	8.0087	0.8452
5	6.00	20.00	4.8173	0.8463
6	4.80	15.20	3.6703	0.8456
7	3.25	11.95	2.3668	0.8507
8	2.50	9.45	1.7339	0.8585
9	2.10	7.35	1.3871	0.8661
10	1.70	5.65	1.0694	0.8748
11	1.25	4.40	0.7489	0.8884
12	1.05	3.35	0.5991	0.9040
13	0.90	2.45	0.4891	0.9214
14	0.80	1.65	0.4140	0.9398
15	0.70	0.95	0.3450	0.9588
16	0.60	0.35	0.2817	0.9759
17	0.35	0.00	0.1565	1.0000

Notes:

(2)	Based on industry data.
(3)	For year (t) = 100.00 - cumulative (2) up to year (t).
(4)	(2)/[(1.0 + .05)^((1) - 0.5)]
(5)	Year (t) = [Sum (4), Year (t+1) to Year (17)]/(3) x [(1+.05)^(1)].

Example One Exhibit 1

Selection of Increased Limits Factor (000s) Sheet 6

Accident Year	Losses (000's)	Claims over $100K Counts	Claims over $100K (000's)	Implied ILF
(1)	(2)	(3)	(4)	(5)
1999	$3,100	3	$750	1.170
2000	2,850	2	400	1.075
2001	3,500	1	600	1.167
2002	3,300	0	0	1.000
2003	2,600	1	175	1.030
Total	$15,350	7	$1,925	1.087

(6) Industry ILF		
	a) 100/250	1.280
	b) 100/500	1.540
	c) 100/750	1.710
(7) Selected ILF		
	a) 100/250	1.250
	b) 100/500	1.510
	c) 100/750	1.690

Notes:

(2) - (4)	From Exhibit 1, Sheet 7.
(5)	(2)/[(2) - (4) + (3) x 100].
(6)	Based on industry data.
(7)	Selected judgmentally.

Example One
Summary of Basic Data evaluated as of December 31, 2003

Accident Year	Paid Losses (000's)	O/S Losses (000's)	Incurred Losses (000's)	Reported Counts	Closed Counts	Claims Over $100K: Counts	Claims Over $100K: Incurred Loss (000's)	Claims Over $100K: Paid Loss (000's)	Losses Limited to $100K: Incurred (000's)	Losses Limited to $100K: Paid (000's)	Payroll (00s)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1999	$2,600	$500	$3,100	600	575	3	$750	$400	$2,650	$2,200	$1,400,000
2000	2,200	650	2,850	625	550	2	400	75	2,650	2,125	1,425,000
2001	2,500	1,000	3,500	700	600	1	600	150	3,000	2,350	1,480,000
2002	1,800	1,500	3,300	725	550	0	0	0	3,300	1,800	1,500,000
2003	800	1,800	2,600	725	400	1	175	0	2,525	800	1,525,000
Total	$9,900	$5,450	$15,350	3,375	2,675	7	$1,925	$625	$14,125	$9,275	$7,330,000

Notes:

(2) - (9),(12)	Provided by Company. (8) reflects total value of claims, while (9) reflects the portion of the claim over $100K.
(10)	(4) - (8) + (7) x 100.
(11)	(2) - (9).

Example One
Projected Expenses (000s)

Exhibit 1
Sheet 8

Operating Expense	Projected Cost
(1)	(2)
Reinsurance - WC only	$200
Risk Management Services	150
Accounting Services	40
Actuarial	25
Consultants	5
Legal Services	5
Trust and Bank Fees	1
State Assessments	20
Miscellaneous	5
Total	$451
Subtotal excluding reinsurance	$251
Coverage expenses - reinsurance plus 10% of program expenses	$225

Notes:
(1),(2) Provided by Company.

Example Two
Allocation of Premium (000's)

Exhibit 2
Sheet 1

1. Projected 2005 Premium, $1,000,000 per occurrence limits - discounted funding, 75% confidence level $7,561

Practice	2005 Exposures	Percentage of Exposures	1999-2003 Reported Counts	Percentage of Counts	1999-2003 Incurred Loss (000's)	Percentage of Inc. Loss	Selected Allocation Percentage	Allocated Premium
(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Practice A	105	35.00%	50	37.31%	$3,000	21.83%	31.38%	$3,559
Practice B	75	25.00%	34	25.37%	4,670	33.99%	28.12%	3,189
Practice C	50	16.67%	31	23.13%	1,695	12.34%	17.38%	1,971
Practice D	70	23.33%	19	14.18%	4,375	31.84%	23.12%	2,622
Total	300		134		$13,740			$11,342

Notes:	
(1)	From Exhibit 2, Sheet 2.
(3),(5),(7)	From Exhibit 2, Sheet 5.
(4)	(3)/(3), total.
(6)	(5)/(5), total.
(8)	(7)/(7), total.
(9)	Equal weighting of (4),(6), and (8).
(10)	(1) x (9) x 1.5 (to incorporate an initial capital contribution).

Example Two
Calculation of Discounted Losses - Occurrence Basis

Exhibit 2
Sheet 2

1. Basic Limit Experience Modified Loss Cost (000's)		$22,803			
2. Projected 2005 Exposures		300			
3. Initial Expected Undiscounted Ultimate Incurred Losses (000's)		$6,841			
4. Discount Factor		0.8383			
5. Initial Expected Discounted Ultimate Incurred Losses (000's)		$5,735			
6. First Years Claims Made Factor		0.35			
7. Increased Limit Factors					
a. $500,000 per occurrence		1.500			
b. $1,000,000 per occurrence		1.900			
c. $2,000,000 per occurrence		2.700			
8. Risk Margins at 75% Confidence Level					
a. $500,000 per occurrence		1.250			
b. $1,000,000 per occurrence		1.400			
c. $2,000,000 per occurrence		1.550			
	Expected Level		75% Confidence Level		
	Nominal	Discounted	Nominal	Discounted	
9. Estimated Ultimate Losses (000's)					
a. $500,000 per occurrence	$10,261	$8,602	$12,827	$10,753	
b. $1,000,000 per occurrence	12,998	10,896	18,197	15,254	
c. $2,000,000 per occurrence	18,470	15,484	28,629	24,000	
10. Estimated Expenses (000's)					
a. $500,000 per occurrence	$2,266				
b. $1,000,000 per occurrence	1,466				
c. $2,000,000 per occurrence	766				
11. Profit Loading	10%				
12. Estimated Premium (000's) - First Year Claims Made Basis					
a. $500,000 per occurrence	$6,508	$5,863	$7,506	$6,699	
b. $1,000,000 per occurrence	6,684	5,866	8,705	7,561	
c. $2,000,000 per occurrence	8,034	6,873	11,985	10,184	

Notes:

(1)	From Exhibit 2, Sheet 3.
(2),(6),(7)	Based on industry data.
(3)	(1) x (2)/1000.
(4)	From Exhibit 2, Sheet 6.
(5)	(3) x (4).
(8)	Based on simulation of insured data.
(9)	(3) or (5) multiplied by (7) and/or (8).
(10)	From Exhibit 2, Sheet 7.
(11)	Selected judgmentally.
(12)	[((6) x (9) + (10).]/(1.0 - (11))

Example Two Exhibit 2
Calculation of Basic Limits Loss Costs and Formula Credibility Sheet 3

Accident Year	Basic Limit Losses (000's) (000's)	LDF to Ultimate	Developed Losses (000's)	Base Class Equivalent Exposures	Developed Loss Cost Per Expos. Unit	Basic Limits Industry Expected Loss Cost	Ratio of Actual to Industry Loss Cost	Exposure Weights
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1999	$1,300	1.163	$1,512	280	$5,399	$18,000	0.300	0.295
2000	2,510	1.250	3,138	285	11,009	19,800	0.556	0.279
2001	870	1.538	1,338	272	4,921	21,780	0.226	0.217
2002	1,545	2.381	3,679	300	12,262	23,958	0.512	0.154
2003	1,665	6.667	11,100	300	37,000	26,354	1.404	0.055
Total	$7,890		$20,766	1,437	$14,451			1.000

(10) Weighted Average Ratio	0.449
(11) Credibility	0.350
(12) Credibility Weighted Average	0.807
(13) Selected Ratio	0.750
(14) Industry Loss Cost at 7/04	$30,404
(15) Experience Modified Loss Cost	$22,803

Notes:

(2)	Exhibit 2, Sheet 5, Column (4) - Exhibit 2, Sheet 4, Sum of Column (7) by year.
(3),(7),(14)	Based on industry data.
(4)	(2) x (3).
(5)	From Exhibit 2, Sheet 5.
(6)	(4) x 1000/(5).
(8)	(6)/(7).
(9)	Based on (3) and (5).
(10)	Weighted average of (8), using weights in (9).
(11)	Based on (3) and (5), and a full credibility standard of 40,000.
(12)	(10) x (11) + [1.0 - (11)].
(13)	Selected judgmentally.
(15)	(13) x (14).

Example Two Exhibit 2
Summary of Claims Over $100K (000's) Sheet 4

						Loss Over $100K	
Practice	Acc. Year	Status	Incurred	Paid	O/S	Incurred	Paid
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Practice A	1998	open	$200	$100	$100	$100	$0
Practice A	1998	open	500	50	450	400	0
Practice A	2000	open	100	50	50	0	0
Practice A	2000	open	500	50	450	400	0
Practice B	1999	open	$500	$200	$300	$400	$100
Practice B	1999	open	500	100	400	400	0
Practice B	2001	open	500	50	450	400	0
Practice B	2002	open	500	0	500	400	0
Practice C	2000	closed	$200	$100	$100	$100	$0
Practice C	2000	open	200	175	25	100	75
Practice C	2000	open	150	125	25	50	25
Practice D	1998	open	$200	$100	$100	$100	$0
Practice D	1999	open	125	25	100	25	0
Practice D	2000	open	1,150	100	1,050	1,050	0
Practice D	2000	open	2,000	125	1,875	1,900	25
Practice D	2002	open	125	50	75	25	0
TOTAL			$7,450	$1,400	$6,050	$5,850	$225
Totals by Accident Year							
1998			$900	$250	$650	$600	$0
1999			1,125	325	800	825	100
2000			4,300	725	3,575	3,600	125
2001			500	50	450	400	0
2002			625	50	575	425	0
TOTAL			$7,450	$1,400	$6,050	$5,850	$225

Notes:
(2) - (6) Provided by the Broker.
(7),(8) Excess of (4) and (5) over 100K.

Example Two
Summary of Basic Data evaluated as of December 31, 2003

Exhibit 2
Sheet 5

Accident Year	Losses (000's)	Losses (000's)	Losses (000's)	Reported Counts	Closed Counts	Physicians FTEs	Average Reported	Average O/S	Average Paid	Reported Frequency
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Practice A										
1999	$200	$500	$700	20	8		$35,000	$41,667	$25,000	#N/A
2000	50	60	110	5	4		22,000	60,000	12,500	#N/A
2001	650	700	1,350	10	6	100	135,000	175,000	108,333	0.100
2002	75	450	525	3	2	110	175,000	450,000	37,500	0.027
2003	15	300	315	12	6	105	26,250	50,000	2,500	0.114
Total	$990	$2,010	$3,000	50	26	315	$60,000	$83,750	$38,077	0.159
Practice B										
1999	$0	$100	$100	2	1	75	$50,000	$100,000	$0	0.027
2000	800	1,200	2,000	5	3		400,000	600,000	266,667	#N/A
2001	20	700	720	10	3	60	72,000	100,000	6,667	0.167
2002	400	600	1,000	12	5		83,333	85,714	80,000	#N/A
2003	150	700	850	5	4	75	170,000	700,000	37,500	0.067
Total	$1,370	$3,300	$4,670	34	16	210	$137,353	$183,333	$85,625	0.162
Practice C										
1999	$100	$0	$100	5	1	50	$20,000	$0	$100,000	0.100
2000	75	50	125	7	3	50	17,857	12,500	25,000	0.140
2001	150	850	1,000	8	5	50	125,000	283,333	30,000	0.160
2002	20	300	320	5	4	50	64,000	300,000	5,000	0.100
2003	50	100	150	6	3	50	25,000	33,333	16,667	0.120
Total	$395	$1,300	$1,695	31	16	250	$54,677	$86,667	$24,688	0.124
Practice D										
1999	$700	$300	$1,000	3	2	55	$333,333	$300,000	$350,000	0.055
2000	600	500	1,100	4	2	60	275,000	250,000	300,000	0.067
2001	500	900	1,400	2	1	62	700,000	900,000	500,000	0.032
2002	50	50	100	5	4	65	20,000	50,000	12,500	0.077
2003	25	750	775	5	1	70	155,000	187,500	25,000	0.071
Total	$1,875	$2,500	$4,375	19	10	312	$230,263	$277,778	$187,500	0.061
TOTAL										
1999	$1,000	$900	$1,900	30	12	280	$63,333	$50,000	$83,333	0.107
2000	1,525	1,810	3,335	21	12	285	158,810	201,111	127,083	0.074
2001	1,320	3,150	4,470	30	15	272	149,000	210,000	88,000	0.110
2002	545	1,400	1,945	25	15	300	77,800	140,000	36,333	0.083
2003	240	1,850	2,090	28	14	300	74,643	132,143	17,143	0.093
Total	$4,630	$9,110	$13,740	134	68	1,437	$102,537	$138,030	$68,088	0.093

Notes:

(2) - (7)	Provided by the Broker. Total (7) includes estimates for missing periods. (7) is on a base class equivalent basis.
(8)	(4) x 1000/(5).
(9)	(3) x 1000/[(5) - (6)].
(10)	(2) x 1000/(6).
(11)	(5)/(7).

Example Two
Calculation of Discount Factor
Assumed Discount Rate of 5%

Exhibit 2
Sheet 6

Year (t)	Percent Paid in Year (t)	Percent Unpaid at end of Year (t)	Present Value of Payments in Year (t)	Avg. Disc. Factor for Remaining Payments at end of Year (t)
(1)	(2)	(3)	(4)	(5)
0		100.00		0.8383
1	5.00	95.00	4.8795	0.8726
2	12.00	83.00	11.1531	0.9005
3	26.00	57.00	23.0144	0.9095
4	22.00	35.00	18.5464	0.9111
5	14.00	21.00	11.2403	0.9113
6	9.00	12.00	6.8818	0.9061
7	4.00	8.00	2.9129	0.9147
8	3.00	5.00	2.0807	0.9219
9	2.00	3.00	1.3211	0.9302
10	1.00	2.00	0.6291	0.9527
11	1.00	1.00	0.5991	0.9759
12	1.00	0.00	0.5706	1.0000

Notes:

(2)	Based on industry data.
(3)	For year (t) = 100.00 - cumulative (2) up to year (t).
(4)	(2)/[(1.0 + .05)^((1) - 0.5)]
(5)	Year (t) = [Sum (4), Year (t+1) to Year (12)]/(3) x [(1+.05)^(1)].

Example Two
Projected Expenses (000s)

Exhibit 2
Sheet 7

Operating Expense (1)	Projected Cost (2)
Reinsurance	
a. $500,000 per occurrence	$2,000
b. $1,000,000 per occurrence	1,200
c. $2,000,000 per occurrence	500
Risk Management Services	$150
Accounting Services	40
Actuarial	25
Consultants	5
Legal Services	5
Trust and Bank Fees	1
Bonds/D&O Insurance	20
Annual Meeting	10
State and Federal Taxes	5
Miscellaneous	5
Total	
a. $500,000 per occurrence	$2,266
b. $1,000,000 per occurrence	1,466
c. $2,000,000 per occurrence	766

Notes:
(1),(2) Provided by Broker.

Example Three Exhibit 3
Calculation of Indicated Funding Sheet 1

1. Estimated Projected Claims		7	
2. Estimated Average Severity		$159,734	
3. Estimated Ultimate Losses		$1,118,100	
4. Risk Margin at			
a. 75% Confidence Level		1.350	
b. 90% Confidence Level		1.700	
c. 95% Confidence Level		2.250	
5. Discount Factor @ 5%		0.866	
6. Expenses		$20,000	
7. Indicated Funding at	Nominal		Discounted
a. Expected Level	$1,138,100		$988,318
b. 75% Confidence Level	1,529,435		1,327,230
c. 90% Confidence Level	1,920,770		1,666,141
d. 95% Confidence Level	2,535,725		2,198,716

Notes:
(1) From Exhibit 3, Sheet 3.
(2) From Exhibit 3, Sheet 2.
(3) (1) x (2).
(4) Based on simulation of healthcare system experience.
(5) From Exhibit 3, Sheet 4.
(6) Provided by the healthcare system.
(7) (3) x (4) (for higher confidence levels) x (5) (for discounted results) + (6).

Example Three
Calculation of Severity

Exhibit 3
Sheet 2

		Probability
A. Outcome A - Fatal Claim		5%
1. Estimated Lost Wages	$1,333	
2. Estimated Medical Costs	500,000	
3. Estimated Survivor Benefits	1,032,307	
4. Total	$1,533,640	
B. Outcome B - Permanent Total Claim		10%
1. Estimated Lost Wages	$416,212	
2. Estimated Medical Costs	100,000	
3. Estimated Future Medical Costs	141,471	
4. Total	$657,683	
C. Outcome C - Eight Week Injury		85%
1. Estimated Lost Wages	$5,333	
2. Estimated Medical Costs	15,000	
3. Total	20,333	
D. Combined Severity (weighted average of A-C)	$159,734	

Notes:

All 3 outcomes assume injured worker currently earns 1,000 per week, a 2/3 replacement rate, 7.5% annual future medical inflation and 4% annual COLA adjustment.
Outcome A: Assumes 2 weeks of wage loss prior to death and 20 years of survivor benefits.
Outcome B: Assumes 10 years of lost wages, annual medical costs of 10,000 in current dollars.
Outcome C: Assumes 8 weeks lost wages.
Combined severity based on probability weighting of severity by outcome.
Probability of each outcome based on industry data and healthcare system input.

Example Three Exhibit 3
Projection of Claim Frequency Sheet 3

1. Estimated Employees	
a. Estimated Payroll (00's)	$4,000,000
b. Average Salary	50,000
c. Estimated Headcount	8,000
<u>A. Direct Exposure (for vaccinated workers)</u>	
2. Estimated % of Workers Vaccinated	1.50%
3. Estimated Number of Vaccinated Workers	120
4. Estimated Percentage of Vaccinated Workers Contracting Smallpox	2.00%
5. Estimated Number of Vaccinated Workers Contracting Smallpox	2
<u>B. Indirect Exposure (non-vaccinated workers exposed by vaccinated workers)</u>	
6. Estimated Percentage of Non-vaccinated Workers exposed to Vaccinated Workers	5.00%
7. Estimated Percentage of Non-vaccinated Workers Contracting Smallpox	1.00%
8. Interaction Effect	1.20
9. Estimated Number of Non-Vaccinated Workers Contracting Smallpox	5
10. Total Projected Claims	7

Notes:

(1a),(1b),(2)	Provided by the healthcare system.
(1c)	(1a) x (1b).
(3)	(1c) x (2).
(4),(6),(7)	Based on industry information and input from healthcare system.
(5)	(3) x (4).
(8)	Estimated based on healthcare system input.
(9)	(1c) x [1.0 - (2)] x (6) x (7) x (8).
(10)	(5) + (9).

Example Three Exhibit 3
Calculation of Discount Factor Sheet 4
Assumed Discount Rate of 5%

Year (t)	Percent Paid in Year (t)	Percent Unpaid at end of Year (t)	Present Value of Payments in Year (t)	Avg. Disc. Factor for Remaining Payments at end of Year (t)
(1)	(2)	(3)	(4)	(5)
0		100.00		0.8660
1	63.63	36.37	62.0988	0.7075
2	2.33	34.04	2.1657	0.7236
3	2.44	31.60	2.1570	0.7393
4	2.55	29.05	2.1488	0.7545
5	2.67	26.39	2.1408	0.7687
6	2.79	23.60	2.1333	0.7815
7	2.92	20.68	2.1261	0.7917
8	3.06	17.62	2.1193	0.7978
9	3.20	14.42	2.1129	0.7962
10	3.35	11.07	2.1069	0.7789
11	0.92	10.15	0.5525	0.7990
12	0.96	9.19	0.5472	0.8196
13	1.00	8.19	0.5420	0.8406
14	1.04	7.16	0.5369	0.8620
15	1.08	6.08	0.5318	0.8839
16	1.12	4.96	0.5267	0.9062
17	1.17	3.79	0.5217	0.9290
18	1.21	2.57	0.5167	0.9522
19	1.26	1.31	0.5118	0.9759
20	1.31	0.00	0.5069	1.0000

Notes:

(2)	Based on industry data.
(3)	For year (t) = 100.00 - cumulative (2) up to year (t).
(4)	(2)/[(1.0 + .05)^((1) - 0.5)]
(5)	Year (t) = [Sum (4), Year (t+1) to Year (17)]/(3) x [(1+.05)^(1)].

Generalized Minimum Bias Models

Luyang Fu, Ph. D. and Cheng-sheng Peter Wu, FCAS, ASA, MAAA

Abstract:
In this research, we propose a flexible and comprehensive approach for minimum bias models -- "Generalized Minimum Bias Models"(GMBM). Unlike the Generalized Linear Models (GLMs) that require the exponential family distribution assumption of response variables, the GMBM approach relaxes the distribution assumption. In addition, due to its model selection flexibility, we believe that GMBM will improve the accuracy and the goodness of fit of classification rates. All the multiplicative minimum bias models published to date and the commonly used multiplicative GLMs (such as Gamma, Poisson, normal, inverse Gaussian) can be proved as special cases of GMBM.

Keywords: GMBM, GLMs, Classification Ratemaking, Weighted Average.

1. INTRODUCTION

Minimum bias models have had a long history for property and casualty actuaries. Until recent interest in generalized linear models (GLM), minimum bias approach was the major technique used by actuaries in determining the rate relativities for a multiple rating variables class plan. Numerous studies have shown that these two related multivariate procedures can reduce the estimation errors from one-way analysis.

Bailey and Simon (1960) originally considered the biases in the classification ratemaking and introduced the minimum bias models. Bailey (1963) summarized the minimum bias theory and proposed two iterative methods (one multiplicative and one additive), which later became popular with the property and casualty actuaries. Because multiplicative models are more popular than additive ones, the following sections will focus on multiplicative models, and the discussion for the additive models will be given in the appendix.

Let $r_{i,j}$ and $w_{i,j}$ be the observed relativity and weight (earned exposure or number of claims) for the classification i and j, respectively; and x_i and y_j be the relativities for the classification i and classification j, respectively. The multiplicative formula proposed by Bailey (1963) is:

$$\text{Model 1:} \quad \hat{x}_i = \frac{\sum_j w_{i,j} r_{i,j}}{\sum_j w_{i,j} y_j} \qquad (1).$$

where i=1,2, to m; and j=1,2, to n. Similarly, $\hat{y}_j = \frac{\sum_i w_{i,j} r_{i,j}}{\sum_i w_{i,j} x_i}$.

Model 1 can be derived by the maximum likelihood (ML) method assuming Poisson distribution, and is also called "ML Poisson model".

Brown (1988) expanded the minimum bias method by using additional types of bias functions. He linked the minimum bias method to statistical theories by maximizing the likelihood functions to calculate the parameter relativity, and introduced four more minimum bias models (three multiplicative and one additive). In addition to the Poisson model in Equation (1), the ML exponential model is:

$$\text{Model 2:} \quad \hat{x}_i = \frac{1}{n}\sum_j \frac{r_{i,j}}{y_j} \qquad (2);$$

and the ML normal model is:

$$\text{Model 3:} \quad \hat{x}_i = \frac{\sum_j w_{i,j}^2 r_{i,j} y_j}{\sum_j w_{i,j}^2 y_j^2} \qquad (3);$$

and the least-square multiplicative model is:

$$\text{Model 4:} \quad \hat{x}_i = \frac{\sum_j w_{i,j} r_{i,j} y_j}{\sum_j w_{i,j} y_j^2} \qquad (4).$$

The formats of Models 1-4 are simple and straightforward. So, compared to GLM, one main advantage of the minimum bias approach is that it is easy to understand and easy to use.

Another minimum bias model by Bailey and Simon (1960) has a relatively complicated format:

$$\text{Model 5:} \quad \hat{x}_i = \left(\frac{\sum_j w_{i,j} r_{i,j}^2 y_j^{-1}}{\sum_j w_{i,j} y_j} \right)^{1/2} \qquad (5).$$

Feldblum and Brosius (2002) summarized these minimum bias models into four categories: "balance principle", "least squares", "χ-squared", and "maximum likelihood":

- Model 1 could be derived from the so-called "balance principle", that is, *"the sum of the indicated relativity = the sum of observed relativity"*. Such balance relationship is:

  $$\sum_j w_{i,j} r_{i,j} = \sum_j w_{i,j} x_i y_j \ .$$

- Model 4 can be derived by minimizing the sum of square-error:

  $$\underset{x,y}{Min} \sum_{i,j} w_{i,j} (r_{i,j} - x_i y_j)^2 \ .$$

- Models 1, 2, and 3 can be derived from the associated log likelihood functions of observed loss (or relativity).

- Model 5 can be derived by minimizing the "χ-squared" error, the square error divided by the indicated relativity:

  $$\underset{x,y}{Min} \sum_{i,j} w_{i,j} \frac{(r_{i,j} - x_i y_j)^2}{x_i y_j} \ .$$

Mildenhall (1999) in his milestone paper further demonstrated that classification rates determined by various linear bias functions are essentially the same as those from GLM models. One main advantage of using statistical models such as GLM is that the characteristics of the models, such as parameters' confidence intervals and hypothesis testing, can be thoroughly studied and determined by statistical theories. Another advantage is that GLM models may be more efficient because they do not require actuaries to program the iterative process in determining the parameters[1]. However, this advantage can be discounted due to the powerful calculation capability associated with modern computers. Due to these advantages, GLMs are becoming more popular in recent years. Of course, actuaries need to acquire the necessary statistical knowledge in understanding and applying the GLM models.

One issue associated with most previous works on the minimum bias models and GLM is the model-selection limitation. GLMs assume the underlying distributions are from the

---

[1] GLMs may also involve iterative approach. The most commonly used numerical method to solve the GLM is the "iterative reweighted least square" algorithm. The discussion of calculation efficiency is given in the appendix.

exponential family, such as Poisson, Gamma, normal, negative binomial, and inverse Gaussian. On the other hand, only five types of multiplicative models and four types of additive models are available from previous minimum bias work[2]. These limitations, we believe, may reduce estimation accuracy in practice since insurance and actuarial data are rarely perfect and may not fit the exponential family of distributions or existing bias models well.

It is with this motivation that in this study, we propose a more flexible and comprehensive approach within the minimum bias framework, called "Generalized Minimum Bias Model". The key features of GMBM are:

- It does not assume a specific form of distributions, which increases the application appropriateness and model-selection flexibility.
- Due to its flexibility, it will improve the accuracy and the goodness of fit of classification rates. We will show the empirical evidence later.
- Similar to past minimum bias models, it is easy to understand and does not require advanced statistical knowledge.
- While GMBM still requires the iterative process in determining the parameters, we believe that the effort is not significant with today's powerful modern computers. We will prove in the appendix that the iterative process required to calculate the GMBM parameters can converge rapidly.

All five existing multiplicative models are proved to be the special cases of GMBM. In addition, we will show several more bias models that actuaries may consider for ratemaking based on insurance data.

The numerical analysis given later is based on severity data for private passenger auto collision given in Mildenhall (1999) and McCullagh and Nelder (1989). The results for selected generalized minimum bias models will be compared to those from the GLM models. Following Bailey and Simon (1960), the weighted absolute bias and the Pearson Chi-Squared statistic are used to measure the goodness of fit. We also calculate the weighted absolute percentage bias, which indicates how much the errors are relative to the predicted values. The empirical results indicate that actuaries can improve the accuracy of classification rates by using the appropriate generalized minimum bias models.

---

[2] Feldblum and Brosius (2002) listed six multiplicative minimum bias models in their summary table. However, the balance principle model is the same as the maximum likelihood Poisson model.

The paper is organized as follows:

- Section 2 discusses the details of 2-parameter and 3-parameter GMBM models.
- Section 3 reviews numerical results for a severity case study.
- Section 4 outlines our conclusions.
- More details and insights for the statistical theories associated with GMBM will be given in Appendix 1.
  - Appendix 1.1 analyzes the bias function of GMBM, proves that GLM with log link are special cases of GMBM, and explores GMBM from the perspective of generalized balance principle.
  - Appendix 1.2 discusses the relativity link functions of GMBM, addresses the difference in the link functions between GLM and GMBM, and analyzes GMBM from the perspective of maximum likelihood method.
  - Appendix 1.3 investigates the possibility of further generalizing GMBM models, explores GMBM from the perspective of deviance functions.
  - Appendix 1.4 explores the additive models for GMBM.
  - Appendix 1.5 discusses the calculation efficiency of GMBM. It shows that GMBM could converge rapidly and is not necessarily inefficient in numerical calculations.
- Appendix 2 reports the numerical results for the severity example discussed in Section 3 with several selected GMBM models.
- Appendix 3 shows the numerical iterative results in Appendix 1.5 for selected GMBM and GLM models.

2. GENERALIZED MINIMUM BIAS MODELS - GMBM

2-Parameter GMBM

Following the notation used previously, in the multiplicative framework for two rating factors, the expected relativity for cell (i,j) should be equal to the product of x_i and y_j:

$$E(r_{i,j}) = \mu_{i,j} = x_i y_j \qquad (6).$$

By (6), there are a total of n alternative estimates for x_i and a total of m estimates for y_j:

$$\begin{aligned} \hat{x}_{i,j} &= r_{i,j} / y_j, \quad j = 1, \quad 2, \quad to \quad n \\ \hat{y}_{j,i} &= r_{i,j} / x_i, \quad i = 1, \quad 2, \quad to \quad m, \end{aligned} \qquad (7).$$

Following actuarial convention, the final estimates of x_i and y_j could be calculated by the weighted average of $\hat{x}_{i,j}$ and $\hat{y}_{j,i}$. If we use the straight average to estimate the relativity:

$$\hat{x}_i = \sum_j \frac{1}{n}\hat{x}_{i,j} = \frac{1}{n}\sum_j \frac{r_{i,j}}{y_j} \qquad (8).$$

Similarly, $\hat{y}_j = \sum_i \frac{1}{m}\hat{y}_{j,i} = \frac{1}{m}\sum_i \frac{r_{i,j}}{x_i}$. This is Model 2, the ML exponential model introduced by Brown (1988).

If the relativity-adjusted number of claims, $w_{i,j}x_i$ or $w_{i,j}y_j$, is used as the weight in determining the estimates:

$$\hat{x}_i = \sum_j \frac{w_{i,j}y_j}{\sum_j w_{i,j}y_j}\hat{x}_{i,j} = \sum_j \frac{w_{i,j}y_j}{\sum_j w_{i,j}y_j}\frac{r_{i,j}}{y_j} = \frac{\sum_j w_{i,j}r_{i,j}}{\sum_j w_{i,j}y_j} \qquad (9).$$

Similarly, $\hat{y}_j = \sum_i \frac{w_{i,j}x_i}{\sum_i w_{i,j}x_i}\hat{y}_{i,j} = \sum_i \frac{w_{i,j}x_i}{\sum_i w_{i,j}x_i}\frac{r_{i,j}}{x_i} = \frac{\sum_i w_{i,j}r_{i,j}}{\sum_i w_{i,j}x_i}$. The resulting model is the same as Model 1, the "balance principle" or ML Poisson model.

If the square of the relativity-adjusted number of claims, $w_{i,j}^2x_i^2$ or $w_{i,j}^2y_j^2$, is used as the weight:

$$\hat{x}_i = \sum_j \frac{w_{i,j}^2y_j^2}{\sum_j w_{i,j}^2y_j^2}\hat{x}_{i,j} = \frac{\sum_j w_{i,j}^2r_{i,j}y_j}{\sum_j w_{i,j}^2y_j^2} \qquad (10).$$

The resulting model is the same as Model 3, the ML normal model.

If the number of claims adjusted by the square of relativity, $w_{i,j}x_i^2$ or $w_{i,j}y_j^2$, is used as the weights:

$$\hat{x}_i = \sum_j \frac{w_{i,j} y_j^2}{\sum_j w_{i,j} y_j^2} \hat{x}_{i,j} = \frac{\sum_j w_{i,j} r_{i,j} y_j}{\sum_j w_{i,j} y_j^2} \qquad (11).$$

The resulting model is the same as Model 4, the least-square model.

From the above results, we propose the 2-parameter generalized minimum bias approach by using $w_{i,j}^p x_i^q$ and $w_{i,j}^p y_j^q$ as the weights for the bias function:

$$\text{2-Parameter GMBM: } \hat{x}_i = \sum_j \frac{w_{i,j}^p y_j^q}{\sum_j w_{i,j}^p y_j^q} \hat{x}_{i,j} = \frac{\sum_j w_{i,j}^p r_{i,j} y_j^{q-1}}{\sum_j w_{i,j}^p y_j^q} \quad (12).$$

When,

- p=q=0, it is the ML exponential model, Model 2;
- p=q=1, it is the ML Poisson model, Model 1;
- p=q=2, it is the ML normal model, Model 3
- p=1 and q=2, it is the least-square model, Model 4.

In addition, there are two more models that correspond to GLM with the exponential family of Gamma and inverse Gaussian distributions[3]. When the number of claims is used as the weights, that is, p=1 and q=0, the GMBM model is a GLM Gamma model and becomes:

$$\text{Model 6: } \hat{x}_i = \sum_j \frac{w_{i,j}}{\sum_j w_{i,j}} \hat{x}_{i,j} = \frac{\sum_j w_{i,j} r_{i,j} y_j^{-1}}{\sum_j w_{i,j}} \qquad (13).$$

When p=1 and q=-1, the GMBM model is a GLM Inverse Gaussian model and becomes:

$$\text{Model 7: } \hat{x}_i = \sum_j \frac{w_{i,j} y_j^{-1}}{\sum_j w_{i,j} y_j^{-1}} \hat{x}_{i,j} = \frac{\sum_j w_{i,j} r_{i,j} / y_j^2}{\sum_j w_{i,j} / y_j} \qquad (14).$$

[3] For detailed information, please refer to Section 7 of Mildenhall (1999).

Equation (12) indicates that in theory there is no limitation for the values of p and q to be used. It is with this feature that GMBM should greatly enhance the flexibility for actuaries when they apply the models to fit their data. Of course, in reality we do not expect that extreme values for p and q will be found useful. In ratemaking applications, earned premium could be used if exposure is not available. Normalized premium (premium divided by relativity) is a reasonable option for the weight. This suggests that q could be negative. In general, p should be positive: the more claims/exposure/premium, the more weight assigned. In the numerical analysis given later, we will test the model with the p and q values ranging from 0 to 2.

In actuarial exercises, we often exclude the extremely high and low values from the weighted average to yield more robust results. In case of several rating variables, there may be thousands of alternative estimates. Actuaries have the flexibility to use the weighted average within selected ranges (e.g. the average without the highest and the lowest 1% percentile). This is similar to concept of "trimmed" regression used with GLMs whereby observations with undue influence on fitted value are removed.

3-Parameter GMBM

So far, we have used the 2-parameter GMBM in Equation (12) to represent several commonly used minimum bias models, Models 1 to 4, but not Model 5, the "χ-squared' multiplicative model. In order to represent Model 5, we further expand the 2-parameter GMBM to a 3-parameter GMBM using the link function concept from GLM.

One generalization of GLMs compared to linear model is to introduce a link function to link the linear predictor to the response variable. Similarly, we introduce a relativity link function, f, which links the minimum bias estimate to the relativity. Of course, this relativity link function is different in several aspects from the link function in GLMs. In GLMs, the link function determines the type of model: log link implies a multiplicative model and identity link implies an additive model. This is not the case for GMBM. A multiplicative GMBM model, for example, could have a log, power, or exponential link function. The detailed discussion of GMBM link function and its difference from GLM link function will be given in Appendix 1.2.

For a 3-parameter GMBM model, instead of using (7), we estimate the relativity link functions of $f(\hat{x}_i)$ and $f(\hat{y}_j)$ from $f(\hat{x}_{i,j})$ and $f(\hat{y}_{j,i})$ first; and then calculate $\hat{x}_i$ and

$\hat{y}_j$ by inverting the relativity function, $f^{-1}(f(\hat{x}_i))$ and $f^{-1}(f(\hat{y}_j))$. The functions $f(\hat{x}_{i,j})$ and $f(\hat{y}_{j,i})$ can be estimated by:

$$\begin{aligned} f(\hat{x}_{i,j}) &= f(r_{i,j}/y_j), \quad j = 1, \quad 2, \quad to \quad n; \\ f(\hat{y}_{j,i}) &= f(r_{i,j}/x_i), \quad i = 1, \quad 2, \quad to \quad m \end{aligned} \qquad (15).$$

Taking the weighted average using parameters p and q:

$$\begin{aligned} f(\hat{x}_i) &= \sum_j \frac{w_{i,j}^p y_j^q}{\sum_j w_{i,j}^p y_j^q} f(\hat{x}_{i,j}) = \frac{\sum_j w_{i,j}^p y_j^q f(\frac{r_{i,j}}{y_j})}{\sum_j w_{i,j}^p y_j^q} \\ f(\hat{y}_j) &= \sum_i \frac{w_{i,j}^p x_i^q}{\sum_i w_{i,j}^p x_i^q} f(\hat{y}_{j,i}) = \frac{\sum_i w_{i,j}^p x_i^q f(\frac{r_{i,j}}{x_i})}{\sum_i w_{i,j}^p x_i^q} \end{aligned} \qquad (16).$$

Thus,

$$\begin{aligned} \hat{x}_i &= f^{-1}\left(\frac{\sum_j w_{i,j}^p y_j^q f(\frac{r_{i,j}}{y_j})}{\sum_j w_{i,j}^p y_j^q} \right) \\ \hat{y}_j &= f^{-1}\left(\frac{\sum_i w_{i,j}^p x_i^q f(\frac{r_{i,j}}{x_i})}{\sum_i w_{i,j}^p x_i^q} \right) \end{aligned} \qquad (17).$$

One possible selection of the relativity link function is the power function, $f(\hat{x}_i) = \hat{x}_i^k$ and $f(\hat{y}_j) = \hat{y}_j^k$. In this case, equation (17) becomes a 3-parameter GMBM:

$$\text{3-Parameter GMBM: } \hat{x}_i = \left(\frac{\sum_j w_{i,j}^p r_{i,j}^k y_j^{q-k}}{\sum_j w_{i,j}^p y_j^q} \right)^{1/k} \qquad (18).$$

When k=2, p=1, and q=1, Equation (18) is equivalent to:

$$\hat{x}_i = \left(\frac{\sum_j w_{i,j} r_{i,j}^2 y_j^{-1}}{\sum_j w_{i,j} y_j} \right)^{1/2} \qquad (19),$$

and this is Model 5, the "χ-squared' multiplicative model.

Another example of a new iterative model is for k=1/2, p=1, and q=1:

$$\text{Model 8:} \quad \hat{x}_i = \left(\frac{\sum_j w_{i,j} r_{i,j}^{1/2} y_j^{1/2}}{\sum_j w_{i,j} y_j} \right)^2 \qquad (20).$$

In the numerical analysis given next, we will test a series of models with the value of k ranging from 0.5 to 3.

3. NUMERICAL ANALYSIS WITH A SEVERITY CASE STUDY

The numerical analysis is based on the severity data for private passenger auto collision given in Mildenhall (1999) and McCullagh and Nelder (1989). It includes thirty-two severity observations for two classification variables: eight age groups and four types of vehicle-use. The GMBMs are tested at hundreds of cases: k at 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0; and p at 0, 0.5, 1.0, 1.5, and 2.0; and q varies from –2.5 to 4.0. The results from multiplicative GLMs with Poisson, Gamma, and inverse Gaussian distributions are compared to those from GMBMs.

Four criteria are used to evaluate the performance of GMBMs: the absolute bias, the absolute percentage bias, the Pearson Chi-Squared statistic, and the combination of absolute bias and the Chi-Squared statistic:

- The weighted absolute bias (wab) criterion is proposed by Bailey and Simon (1960). It is the weighted average of absolute dollar difference between the observations and fitted values:

$$wab = \frac{\sum w_{i,j} \mid r_{i,j} - x_i y_j \mid}{\sum w_{i,j}}$$

- The second one, weighted absolute percentage bias ($wapb$), measures the absolute bias relative to the predicted values:

$$wapb = \frac{\sum w_{i,j} \frac{|r_{i,j} - x_i y_j|}{x_i y_j}}{\sum w_{i,j}}$$

- The weighted Pearson Chi-square ($wChi$) statistic is also proposed by Bailey and Simon (1960) and it is appropriate to test "differences between the raw data and the estimated relativities should be small enough to be caused by chance":

$$wChi = \frac{\sum w_{i,j} \frac{(r_{i,j} - x_i y_j)^2}{x_i y_j}}{\sum w_{i,j}}$$

- Lastly, we combine the absolute bias and Pearson Chi-square statistic, $\sqrt{wab * wChi}$, to be the fourth criterion for the model selection.

The numeric results of the study are given in Appendix 2. Tables 1-6 report the GMBM relativities for k=0.5, 1.0, 1.5, 2.0, 2.5, and 3.0, respectively. Table 7 lists the GLM relativities for Gamma, Poisson, and inverse Gaussian distributions. In addition, Table 8 specifically shows GMBM relativities for several selected cases when k=1 and p=1 in Table 2 because these cases are corresponding to the GLM results in Table 7. In all the cases, class "age 60+" and "pleasure" are used as the base. Tables 9-14 show the weighted absolute bias (wab), weighted absolute percentage bias ($wapb$), Chi-square statistic ($wChi$), and combined $\sqrt{wab * wChi}$ for the corresponding GMBM models in Tables 1-6.

The p and q values in Tables 1-6 and Tables 9-14 for each k value are selected by the following rules. For each k, five p values (0, 0.5, 1.0, 1.5, and 2.0) are used; and for each combination of k and p, at least five q values (0, 0.5, 1.0, 1.5, and 2.0) are calculated. If there exists a model with local minimized $\sqrt{wab * wChi}$ among the five values, no further q is used. If $\sqrt{wab * wChi}$ is strictly decreasing (or increasing) with q, we will try additional higher (or lower) q values until the local minimization with respect to $\sqrt{wab * wChi}$ is found. The local optimal model for each combination of k and p is reported in bold and the global optimal model for each k is reported in underlying bold.

From Tables 9-14, we can see that if wab or $wapb$ is used to measure the model performance, for all the k values tested, p=2, and q=0 is the best, and the global minimum error occurs when k=3. The result suggests that the best-fit model, in this example, does not occur with any of the 5 commonly used minimum models of which the underlying distributions are from the exponential family. It clearly demonstrates the fact the insurance data may not be perfect for predetermined distributions. Therefore, the GMBM approach will provide actuaries a more flexible and comprehensive approaching in analyzing their data.

On the other hand, if $wChi$ is used, "χ-squared' model (k=2, p=1, and q=1) provides the best solution. This is expected because "χ-squared' model is calculated by minimizing the Pearson Chi-square statistic.

If we use the criterion of $\sqrt{wab * wChi}$ to select models, k=2.5, p=1, q=-0.5 offers best overall result. The best solution for different k values is different: p=1.5 and q=1 is best when k=0.5 and 1; p=1 and q=-0.5 is best when k=1.5, 2, 2.5, and 3. Again, the 5 commonly used minimum bias models are not the best solution when absolute bias and Chi-square statistic are considered simultaneously.

As stated before, we find that GLMs with common exponential family distribution assumptions are special cases of GMBM (k=1 and p=1). Therefore, the results between Table 7 and Table 8 are the same for the corresponding models:

- when k=1, p=1, and q=2, the "least-square" GMBM has the same results as GLM with normal distribution[4];
- k=1, p=1, and q=1 is the same as Poisson GLM;
- k=1, p=1, and q=0 is the same as Gamma GLM.
- k=1, p=1 and q=-1 is the same as GLM with inverse Gaussian distribution.

We will prove in the appendix that GMBM with k=1 and p=1 ($\hat{x}_i = \dfrac{\sum_j w_{i,j} r_{i,j} y_j^{q-1}}{\sum_j w_{i,j} y_j^{q}}$) is equivalent to the multiplicative GLMs with the variance function of $V(\mu) = \mu^{2-q}$ for assumed exponential family distribution.

It is well known that insurance and actuarial data is generally positively skewed. The skewness for the symmetric normal distribution is zero, and is increasingly positive from Poisson to Gamma to inverse Gaussian distribution. For the GMBM models, the skewness

[4] The underlying assumption of "least-square" regression is that the residuals follow normal distribution. So "least-square" method is same as GLM normal.

can be represented by the difference between p and q (i.e., p-q). The value of p-q is -1 for normal, 0 for Poisson, 1 for Gamma, and 2 for inverse Gaussian. Thus, larger difference between p and q should be selected in GMBM for more skewed data. We expect that GMBM with negative values of p-q will not have good performance in fitting actuarial data.

Based on the result in this research and our experience, we suggest for actuarial applications the following ranges of values for k, p, q:

- k is between 1 and 3.
- p>=q and 0.5<=p<=2.
- The higher the skewness of the data, the larger p-q to use.

4. CONCLUSIONS

In this research, we propose a generalized minimum bias approach by including different weighting functions and relativity link functions in the approach. As indicated by the severity example given previously, insurance and actuarial data are rarely perfect, so we expect that the best fitted results typically will not be based on a predetermined distribution, such as exponential family distributions. Therefore, GMBM can provide actuaries a great deal of flexibility in data fitting and model selection.

In theory, there is no limitation for different weighting functions or relativity link functions to try when GMBM is applied to a dataset. However, due to the fact that insurance and actuarial data is positively skewed in nature, we do not expect that a very wide range of weighting or relativity functions need to be used in practice.

For the severity example used in the study, we tried hundreds of different combinations and identified the best model with the minimum fitted error among the trial models. Two issues may exist for the example. The first issue is that since minimum bias models use an iterative process in determining the parameters, the fact that GMBM further requires multiple models in trial may make the approach even more time consuming and inefficient. However, we do not believe this issue is significant because of the powerful computation capability with modern computers.

Another issue is that the "true" best model with the minimum error may not be one of the models in the trial. This is very possible in practice. Resolving such global minimum error issue requires additional in-depth research and is beyond the scope of this paper.

With the fast development of information technology, people can analyze data in ways they could not imagine a decade ago. Currently there is a strong interest in data mining and predictive modeling in the insurance industry, and this calls for more powerful data analytical tools for actuaries. While some new tools, such as GLM, neural networks, decision Trees, and MARS, have emerged recently and have received a great deal of attention, we believe that the decades-old minimum bias models still have several advantages over other techniques, including easy to understand and easy to use. We hope that our work in improving the flexibility and comprehensiveness for the minimum bias approach is a timely effort and the approach will continue to be a useful tool for actuaries in the future.

5. REERENCES

Bailey, R. A., "Insurance Rates with Minimum Bias", *Proceedings of Casualty Actuarial Society,* Vol. L, Casualty Actuarial Society, (1963).

Bailey, R. A. and L. J., Simon, "Two Studies in Auto Insurance Ratemaking", *Proceedings of Casualty Actuarial Society,* Vol. XLVII, Casualty Actuarial Society, (1960).

Brown, R. L., "Minimum Bias with Generalized Linear Models", *Proceedings of Casualty Actuarial Society,* Vol. LXXV, Casualty Actuarial Society, (1988).

Feldblum, S. and J. E., Brosius, "The Minimum Bias Procedure--A Practitioner's Guide", *Proceedings of Casualty Actuarial Society,* Vol. Fall, Casualty Actuarial Society, (2002).

McCullagh, P., and J. A., Nelder, *Generalized Liner Models*, 2nd Edition, Chapman and Hall, (1989).

Mildenhall, S. J., "A Systematic Relationship Between Minimum Bias and Generalized Linear Models," *Proceedings of Casualty Actuarial Society,* Vol. LXXXVI, Casualty Actuarial Society, (1999).

Venter, G., "Minimum Bias with Generalized Linear Models [Discussion]", *Proceedings of Casualty Actuarial Society,* Vol. LXXVII, Casualty Actuarial Society, (1990).

Appendix 1: Additional Discussions of GMBM Theories

1.1 Bias Functions of GMBM

Bailey and Simon (1960) lay out four criteria for classification relativities. The first criterion is that the rates are balanced for each class and in total. This results in an overall zero ordinary bias, which is the difference between the observed relativity and its fitted value $r_{i,j} - \mu_{i,j}$. Mildenhall (1999) showed that Bailey and Simon's balance criteria is equivalent to a minimum deviance criteria of GLMs of which the bias is measured by linear bias functions; and the bias functions of GLMs are ordinary biases weighted by exposure adjusted by the first order derivative of the GLM link function.

Let Z be the design matrix with rows z_i, h(x) be the inverse function of GLM link function, V(x) be the variance function of the GLM assumed distribution, W be the diagonal matrix of weights with the ith diagonal element of $w_i h'(z_i\beta)/V(\mu_i)$. Mildenhall showed that the bias function of GLM is:

$$Z'W(r-\mu) = 0 \qquad \text{(A.1).}$$ [5]

Let a_i and b_j be the GLM coefficients. A.1 is equivalent to:

$$\begin{aligned} &\sum_{j=1}^{n} \frac{w_{i,j} \dfrac{\partial h(a_i + b_j)}{\partial a_i}(r_{i,j} - \mu_{i,j})}{V(\mu_{i,j})} = 0 \quad for \quad j = 1,2,\ldots,n; \\ &\sum_{i=1}^{m} \frac{w_{i,j} \dfrac{\partial h(a_i + b_j)}{\partial b_j}(r_{i,j} - \mu_{i,j})}{V(\mu_{i,j})} = 0 \quad for \quad i = 1,2,\ldots,m. \end{aligned} \qquad \text{(A.2).}$$

For GLM multiplicative models, h(x) is the exponential function, $h(x) = e^x$, and the most commonly used variance functions are power functions $V(\mu) = \mu^c$. So, A.2 becomes:

$$\sum_{j=1}^{n} w_{i,j}\mu_{i,j}^{1-c}(r_{i,j} - \mu_{i,j}) = 0 \qquad \text{(A.3).}$$

[5] For the detailed explanation, please refer to Section 5 of Mildenhall (1999).

For normal distribution, c=0; for Poisson distribution, c=1, for Gamma distribution, c=2, and for inverse Gaussian distribution, c=3.

From A.3, Mildenhall (1999) further showed that, as an example, the GLM model with log link and Poisson distribution is equivalent to Model 1 $\hat{x}_i = \frac{\sum_j w_{i,j} r_{i,j}}{\sum_j w_{i,j} y_j}$. For more general cases of exponential family distributions with the variance function of $V(\mu) = \mu^c$, A.3 can be transformed into the following to estimate the relativity:

$$\sum_j w_{i,j} \mu_{i,j}^{1-c} (r_{i,j} - \mu_{i,j}) = 0 \Rightarrow x_i^{1-c} \sum_j w_{i,j} y_j^{1-c} (r_{i,j} - x_i\, y_j) = 0$$

$$\Rightarrow \hat{x}_i = \frac{\sum_j w_{i,j} r_{i,j} y_j^{1-c}}{\sum_j w_{i,j} y_j^{2-c}} \qquad \text{(A.4).}$$

Similar to GLM, the bias function of GMBM for a rating variable x can be represented as follows:

$$\sum_j w_{i,j}^P \mu_{i,j}^{q-k} (r_{i,j}^k - \mu_{i,j}^k) = 0 \qquad \text{(A.5).}$$

From A.5 we can derive the 3-parameter GMBM models as given in Equation (17):

$$\sum_j w_{i,j}^P \mu_{i,j}^{q-k} (r_{i,j}^k - \mu_{i,j}^k) = 0 \Rightarrow \sum_j w_{i,j}^P y_j^{q-k} (r_{i,j}^k - x_i^k\, y_j^k) = 0$$

$$\Rightarrow \sum_j w_{i,j}^P y_j^{q-k} r_{i,j}^k - x_i^k \sum_j w_{i,j}^p y_j^q = 0 \Rightarrow \hat{x}_i = \left(\frac{\sum_j w_{i,j}^p r_{i,j}^k y_j^{q-k}}{\sum_j w_{i,j}^p y_j^q} \right)^{1/k} \qquad \text{(A.6).}$$

Comparing A.3 and A.5, we can see that when k=1, the bias for GMBM is measured by the difference between observed and fitted values, and is an average of ordinary bias. Therefore, the GMBM models generalize GLMs on the weights assigned to each ordinary bias since the GLM models for c=0, 1, 2, and 3 can be represented by the GMBM models with the corresponding p and q values. On the other hand, when $k \neq 1$, the GMBM bias is measured by the powered difference of observed and fitted values, which is a further generalization. Compared to GLM, the GMBM models are more general in use in how to measure the bias and how to assign the weights.

Next, we compare A.4 for GLM models and A.6 for GMBM models. With k=1, p=1, and q=2-c, we can see that:

- when c=0, the normal GLM model is the same as the GMBM Model 4 in Equation (4);
- when c=1, the Poisson GLM model is the same as the GMBM Model 1 in Equation (1);
- when c=2, the Gamma GLM model is the same as the GMBM Model 6 in Equation (13);
- when c=3, the Inverse Gaussian GLM model is the same a the GMBM Model 7 in Equation (14).

The GMBM bias function can also be related to a previous minimum bias work from the perspective of "balance principle". Let's rewrite A.4:

$$\sum_j w_{i,j}^P \mu_{i,j}^{q-k}(r_{i,j}^k - \mu_{i,j}^k) = 0 \Rightarrow \sum_j w_{i,j}^P \mu_{i,j}^{q-k} r_{i,j}^k = \sum_j w_{i,j}^P \mu_{i,j}^{q-k} \mu_{i,j}^k \qquad \text{(A.7).}$$

When p=1, q=1, and k=1, A.7 is

$$\sum_j w_{i,j} r_{i,j} = \sum_j w_{i,j} \mu_{i,j} \qquad \text{(A.8).}$$

This is the balance principle by Feldblum and Brosius (2002): "the sum of indicated pure premiums = the sum of the observed loss cost". GMBM further expands in A.7 the balance principle to the "generalized balance principle", that is, "the sum of weighted functions of indicated pure premiums = the sum of weighted functions of the observed loss costs".

1.2 Link Functions of GMBM

In GLMs, the purpose of the link function is to establish the relationship between the predicted value, which is the linear combination of GLM coefficients, and the response variable, such as severity, frequency or pure premium in our typical ratemaking applications. The link functions determine whether the model is multiplicative, additive, or other.

For example, if a GLM with a log link function has coefficients a_i and b_j, then the predicted value is $\hat{Y} = \mathit{Intercept} + a_i + b_j$. The link function then links the predicted value, $\hat{Y}$, to the indicated value, $\mu_{i,j}$, for the response variable**:**

$$\hat{Y} = \log(\mu_{i,j}) \Rightarrow \mu_{i,j} = e^{Intercept} e^{a_i + b_j} \Rightarrow e^{Intercept} x_i y_j \qquad \text{(A.9).}$$

Thus, GLM with a log link function results in a multiplicative model. On the other hand, if the link function is a power function, then

$$\hat{Y} = (\mu_{i,j})^k \Rightarrow \mu_{i,j} = (\text{int}\, ercept + a_i + b_i)^{1/k} \qquad \text{(A.10).}$$

For this case, the GLM model is neither an additive model nor a multiplicative model.

In GMBM, the link function links the iterative coefficients to the classification relativity. Let's assume the power link function to be $A_i = x_i^k$ and $B_j = y_j^k$ for a GMBM model. Then, the relativity is estimated through the following iterative weighted average procedure: $x_i^k = \dfrac{\sum_j w_{i,j}^p r_{i,j}^k y_j^{q-k}}{\sum_j w_{i,j}^p y_j^q}$ and $y_j^k = \dfrac{\sum_i w_{i,j}^p r_{i,j}^k x_i^{q-k}}{\sum_i w_{i,j}^p x_i^q}$, and the GMBM coefficients are directly estimated by inverting the link function:

$$A_i = x_i^k \Rightarrow x_i = A_i^{1/k} = \left(\frac{\sum_j w_{i,j}^p r_{i,j}^k y_j^{q-k}}{\sum_j w_{i,j}^p y_j^q} \right)^{1/k} \qquad \text{(A.11).}$$

If we assume a log link function, the iterative process of GMBM is:

$$A_i = \log(x_i) = \frac{\sum_j w_{i,j}^p y_j^q \log(r_{i,j} / y_j)}{\sum_j w_{i,j}^p y_j^q} \qquad \text{(A.12).}$$

and the GMBM coefficients are estimated by

$$x_i = \exp(A_i) = \exp\left(\frac{\sum_j w_{i,j}^p y_j^q \log(r_{i,j} / y_j)}{\sum_j w_{i,j}^p y_j^q} \right) \qquad \text{(A.13).}$$

One difference between GLM and GMBM for the link function is that in GLM, the link function determines the type of model: log link implies a multiplicative model and identity link implies an additive model. This is not the case for GMBM. A multiplicative GMBM model, for example, could have a log, power, exponential, or any other formats, in theory, for the link function.

Finally, we will show how the GMBM link functions are related to the underlying distribution assumptions in the maximum likelihood procedure. For example, let $L_{i,j}$ be the loss of classification i and j, and B be the base rate. Assuming loss follows a normal distribution, $L_{i,j} = w_{i,j} B r_{i,j}$ and $L_{i,j} \sim N(w_{i,j} B \mu_{i,j}, \sigma^2)$. Following the method in Brown (1980), the density function is:

$$\begin{aligned} f(L_{i,j}) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(L_{i,j} - w_{i,j} B \mu_{i,j})^2 \right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(w_{i,j} B r_{i,j} - w_{i,j} B x_i y_j)^2 \right) \end{aligned} \qquad \text{(A.14).}$$

Minimizing the log likelihood function, we will get:

$$\begin{aligned} l &= \sum_i \sum_j \{-\ln(\sigma\sqrt{2\pi})\} - \frac{B^2}{2\sigma^2} \sum_i \sum_j w_{i,j}^2 (r_{i,j} - x_i y_j)^2 \\ \frac{\partial l}{\partial x_i} &= 0 \Rightarrow x_i = \frac{\sum_j w_{i,j}^2 r_{i,j} y_j}{\sum_j w_{i,j}^2 y_j^2} \end{aligned} \qquad \text{(A.15).}$$

A.15 is the GMBM model with p=2, q=2, and k=1.

Similarly, if we assume the loss square $L_{i,j}^2$ follows a normal distribution, then $L_{i,j}^2 \sim N((w_{i,j} B R_{i,j})^2, \sigma^2)$, where $R_{i,j}^2 = E(r_{i,j}^2)$ and $R_{i,j}^2 = X_i^2 * Y_j^2$
The density function is:

$$\begin{aligned} f(L_{i,j}^2) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(L_{i,j}^2 - w_{i,j}^2 B^2 R_{i,j}^2)^2 \right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(w_{i,j}^2 B^2 r_{i,j}^2 - w_{i,j}^2 B^2 X_i^2 Y_j^2)^2 \right) \end{aligned} \qquad \text{(A.16).}$$

Minimizing the log likelihood function is:

$$l = \sum_i \sum_j \{-\ln(\sigma\sqrt{2\pi})\} - \frac{B^4}{2\sigma^2}\sum_i \sum_j w_{i,j}^4 (r_{i,j}^2 - X_i^2 Y_j^2)^2$$

$$\frac{\partial l}{\partial X_i} = 0 \Rightarrow \frac{\partial l}{\partial X_i^2}\frac{dX_i^2}{dX_i} = 0 \Rightarrow \frac{\partial l}{\partial X_i^2} = 0 \qquad \text{(A.17).}^6$$

$$\Rightarrow X_i^2 = \frac{\sum_j w_{i,j}^4 r_{i,j}^2 Y_j^2}{\sum_j w_{i,j}^4 Y_j^4} \Rightarrow X_i = \left(\frac{\sum_j w_{i,j}^4 r_{i,j}^2 Y_j^2}{\sum_j w_{i,j}^4 Y_j^4}\right)^{1/2}$$

A.17 is the GMBM model with p=4, q=4, and k=2.

Generalizing the maximum likelihood estimation of classification relativity as in A.16 and A.17, we assume that the link function of loss ($L_{i,j}^{k_0}$) follows the same distribution as $L_{i,j}$ in the 2-parameter model with $p = p_0$ and $q = q_0$, then the maximum likelihood estimation of relativities becomes 3-parameter GMBM:

$$\hat{X}_i = \left(\frac{\sum_j w_{i,j}^{p_0 k_0} r_{i,j}^{k_0} Y_j^{(q_0-1)k_0}}{\sum_j w_{i,j}^{p_0 k_0} Y_j^{q_0 k_0}}\right)^{1/k_0} \qquad \text{(A.18).}^7$$

This is the GMBM with $k = k_0$, $p = p_0 k_0$, and $q = q_0 k_0$.

1.3 GMBM based on Deviance Functions

Before Mildenhall (1999), the difference between a measure of bias and a measure of deviance was not discussed. Mildenhall pointed out that the bias could be positive or negative and should be proportional to the difference between the predicted value and the observed value. On the other hand, the deviance is always positive, and is used to measure the goodness of fit. Mathematically, the GLM classification relativities could be obtained by solving zero bias function or minimizing the deviance function. Mildenhall showed that the GLM deviance function with a linear bias function has the following format:

---

[6] Because $X_i Y_j = R_{i,j} \neq \mu_{i,j} = x_i y_j$, X_i and Y_j are biased estimates of x_i and y_j.

[7] Following the derivation for normal distribution, $w_{i,j}^{k_0}$, $r_{i,j}^{k_0}$, $X_i^{k_0}$, and $Y_j^{k_0}$ take the place of $w_{i,j}$, $r_{i,j}$, x_i, and y_j in the likelihood function and solution.

$$d(r,\mu) = 2w\int_{\mu}^{r} \frac{r-t}{V(t)}dt \qquad \text{(A.19).}$$

As given in A.4 and A.6, all of the GMBM models have a generalized linear bias function, $\sum_j w_{i,j}^P y_j^{q-k} r_{i,j}^k - x_i^k \sum_j w_{i,j}^p y_j^q$, and the parameters could be solved by the iterative procedure in minimizing the bias function. However, Mildenhall discussed that the deviance functions may not correspond to linear bias functions (i.e. $d(r,\mu) = w|r-\mu|$). One popular criterion to measure the goodness of fit is the Chi-square statistic, which could also be used as deviance function:

$$d(r,\mu) = \sum_i \sum_j w_{i,j} \frac{(r_{i,j} - \mu_{i,j})^2}{\mu_{i,j}} \qquad \text{(A.20).}$$

Feldblum and Brosius (2002) showed that Model 5 (the "χ-squared" model) could be derived by minimizing the deviance function in A.20.

Following the same generalization work previously for the balance principle, the deviance functions could be also generalized through the weights and the measurement of bias:

$$d(r,\mu) = \sum_i \sum_j w_{i,j}^p \mu_{i,j}^q \frac{(r_{i,j}^k - \mu_{i,j}^k)^2}{\mu_{i,j}^k} \qquad \text{(A.21).}$$

So, A.21 is our proposed generalized deviance function for GMBM, and the classification relativities can be solved through an iterative process by minimizing the generalized deviance function. The iterative procedure starts with setting the first order derivative of A.21 to zero:

$$\frac{\partial d(r,\mu)}{\partial x_i} = 0 \Rightarrow$$

$$x_i^k (q+k) \sum_j w_{i,j}^p y_j^{q+k} + x_i^{-k} (q-k) \sum_j w_{i,j}^p r_{i,j}^{2k} y_j^{q-k} = 2\mathrm{q} \sum_j w_{i,j}^p r_{i,j}^k y_j^q \qquad \text{(A.22).}$$

Because A.22 (or its bias function) does not follow a linear format, the iterative process may not have a simple iterative weighted average formula. Only for certain special values for q, the relativities can be solved by the conventional minimum bias format:

- When q=0,

$$\hat{x}_i = \left(\frac{\sum_j w_{i,j}^p r_{i,j}^{2k} y_j^{-k}}{\sum_j w_{i,j}^p y_j^k} \right)^{\frac{1}{2k}} \quad \text{(A.23).}$$

A special case of A.23 is the "χ-squared' model when k=1.

- When q=k,

$$\hat{x}_i = \left(\frac{\sum_j w_{i,j}^p r_{i,j}^k y_j^k}{\sum_j w_{i,j}^p y_j^{2k}} \right)^{\frac{1}{k}} \quad \text{(A.24).}$$

- When q=-k,

$$\begin{aligned} \hat{x}_i &= \left(\frac{\sum_j w_{i,j}^p r_{i,j}^{2k} y_j^{-2k}}{\sum_j w_{i,j}^p r_{i,j}^k y_j^{-k}} \right)^{\frac{1}{k}} \\ &= \left(\sum_j \frac{w_{i,j}^p r_{i,j}^k y_j^{-k}}{\sum_j w_{i,j}^p r_{i,j}^k y_j^{-k}} (r_{i,j}^k / y_j^k) \right)^{\frac{1}{k}} \end{aligned} \quad \text{(A.25).}$$

In A.25, $w_{i,j}^p r_{i,j}^k y_j^{-k}$ is used as the weight. Compared to $w_{i,j}^p y_j^q$ in Section 2, A.25 adds observed relativity into the weight.

1.4. Additive GMBM Models

Following the same notations given in the text, the expected cost for classification cell (i,j) with an additive model should be equal to the sum of x_i and y_j:

$$E(r_{i,j}) = \mu_{i,j} = x_i + y_j \quad \text{(A.26).}$$

Thus,

$$\begin{aligned} \hat{x}_{i,j} &= r_{i,j} - y_j, \quad j = 1, \quad 2, \quad to \quad n \\ \hat{y}_{j,i} &= r_{i,j} - x_i, \quad i = 1, \quad 2, \quad to \quad m, \end{aligned} \qquad \text{(A.27).}$$

In GMBM multiplicative models, we use the relativity-adjusted claim number, $w_{i,j}^p x_i^q$ and $w_{i,j}^p y_j^q$, as the weighting function and introduce the power link function. However, the weighting functions and the link functions cannot be applied in an additive process.

For the additive GMBM models, we are limited to the following one-parameter model using $w_{i,j}^p$ as the weight:

$$\hat{x}_i = \sum_j \frac{w_{i,j}^p}{\sum_j w_{i,j}^p} \hat{x}_{i,j} = \frac{\sum_j w_{i,j}^p (r_{i,j} - y_j)}{\sum_j w_{i,j}^p} \qquad \text{(A.28).}$$

When p=1, it is the model introduced by Bailey (1963) or the "Balance Principle" model in Feldblum and Brosius (2002). Mildenhall (1999) also proved that it is equivalent to additive normal GLM model. When p=2, it is the ML additive normal model introduced by Brown (1988). When p=0, it is the least squares model by Feldblum and Brosius (2002). There is no further generalization for the additive models with additional parameters or link functions.

We can also extend the model through the deviance function. Because of its additive feature, we can only generalize the "χ-squared" deviance function through the weights:

$$d(r, \mu) = \sum_i \sum_j w_{i,j}^p \frac{(r_{i,j} - \mu_{i,j})^2}{\mu_{i,j}} \qquad \text{(A.29).}$$

Using the first order condition, we solve the numerical results for A.29 with Newton's method as follows:

$$\Delta x_i = \frac{\sum_j w_{i,j}^p \left(\frac{r_{i,j}}{x_i + y_j} \right)^2 - \sum_j w_{i,j}^p}{2 \sum_j w_{i,j}^p \left(\frac{r_{i,j}}{x_i + y_j} \right)^2 \left(\frac{1}{x_i + y_j} \right)} \qquad \text{(A.30).}$$

When p=1, it is the additive model introduced by Bailey and Simon (1960) or the "χ-squared" model by Feldblum and Brosius (2002).

Appendix 1.5: Calculation Efficiency of GMBM

Mildenhall (1999) discussed that one advantage of GLMs compared to minimum bias model is the calculation efficiency because GLMs do not require an iterative process in estimating the parameters. He showed that the additive minimum bias model by Bailey (1963), or GMBM with p=1, does not converge even after 50 iterations[8].

In this study, we propose a simple but improved iterative algorithm. With the improved algorithm we find that GMBM is not necessarily inefficient in numerical calculations. Applying the improved algorithm to the same severity data studied previously in Section 3 and by Mildenhall, most of the GMBM models reported in Appendix 2 converge within 10 iterations. The algorithm could be implemented in major statistical languages, such as SAS, Splus, and Matlab. If the data is not large, it can also be conducted in EXCEL with straightforward VBA codes.

To illustrate the calculation efficiency of GMBM, we first use multiplicative Gamma model because of its popularity in fitting severity data. As for additive model, we will show the iterative process for GMBM with p=1. We will show that the minimum bias model with the proposed algorithm converges in 5 iterations compared to the 50 iterations in Mildenhall (1999). The corresponding GLM models are also solved numerically for comparison using the iterative reweighted least square method[9].

Before showing the detailed results, we would like to discuss the iterative procedure. For the GMBM Gamma model: $x_i = \dfrac{\sum_j w_{i,j} r_{i,j} y_j^{-1}}{\sum_j w_{i,j}}$. In the iterative process, it is better to include as much updated information as possible. Let t be the t-th iteration step,

[8] For detailed information, please refer to Exhibit 5 of Mildenhall (1999).
[9] Iterative reweighted least square method is the most commonly used algorithm for GLMs. The major statistical languages, such SAS and Splus, apply this method to solve GLMs numerically. The discussion of the algorithm is beyond the scope of this paper.

$$
\begin{aligned}
x_{i,t} &= \frac{\sum_j w_{i,j} r_{i,j} y_{j,t-1}^{-1}}{\sum_j w_{i,j}} \\
y_{j,t} &= \frac{\sum_j w_{i,j} r_{i,j} x_{i,t}^{-1}}{\sum_j w_{i,j}}
\end{aligned}
\tag{A.31}
$$

In A.31, $x_{i,t}$ is calculated based on the latest iterative result for $y_{j,t-1}$. When calculating $y_{j,t}$, the latest result of $x_{i,t}$ is used so that $y_{j,t} = \frac{\sum_j w_{i,j} r_{i,j} x_{i,t}^{-1}}{\sum_j w_{i,j}}$. Similarly, for a model that has 3 rating factors, the iterative process is:

$$
\begin{aligned}
x_{i,t} &= \frac{\sum_j w_{i,j,k} r_{i,j,k} y_{j,t-1}^{-1} z_{k,t-1}^{-1}}{\sum_j w_{i,j,k}} \\
y_{j,t} &= \frac{\sum_j w_{i,j,k} r_{i,j,k} x_{i,t}^{-1} z_{k,t-1}^{-1}}{\sum_j w_{i,j,k}} \\
z_{k,t} &= \frac{\sum_j w_{i,j,k} r_{i,j,k} x_{i,t}^{-1} y_{j,t}^{-1}}{\sum_j w_{i,j,k}}
\end{aligned}
\tag{A.32}
$$

In GLMs, a specific classification is usually selected as the base (i.e. age 60+ and pleasure). Mildenhall (1999) also follows this method. We suggest using the average severity as the base in GMBM numerical analysis. Using a specific classification as the base, the numerical value of base varies in each iteration, and the factor of base class has to be forced as one. These may cause additional iterations to solve the functions.

Another well-known issue in numerical analysis is how to set the starting point for t=0. Using average severity as the base, the average factor is one for multiplicative models and average discount is zero for the additive models. Therefore, in this study, we chose the starting values of $x_{i,0}$ and $y_{j,0}$ to be one for the multiplicative models and zero for the additive models.

For both GLMs and GMBM models, the iteration will stop if the errors of factors (or GLM coefficients) are less than 10^{-7} and the errors of dollar values are less than 10^{-4}.

The numerical iterative results are report as follows:

- Table 15 shows the multiplicative factors for the Gamma GMBM using average severity as the base.
- Table 16 transfers those factors using the classification age 60+ and pleasure as the base.
- Table 17 reports the iterative process for the coefficients of GLM with the Gamma distribution and log link.
- Table 18 transfers those coefficients to the multiplicative factors of Gamma GLM.
- Table 19 lists the additive factors for the GMBM with p=1.
- Table 20 shows the additive dollar values for the GMBM with p=1 and uses the classification age 60+ and pleasure as the base.
- Table 21 reports the coefficients of additive normal GLM.

From Tables 15-18, the multiplicative Gamma GMBM converges in 4 iterations. This is as fast as the corresponding GLM model. As expected, the numerical solutions between the two models are identical, and the solutions are also identical to the previous results given in Table 2 for k=1, p=1, and q=0.

Tables 19 and 20 report the iterative process for the GMBM additive model with p=1. Mildenhall (1999) used this model as an example to show that minimum bias model is not efficient. He showed that the minimum bias model still did not converge to the GLM results and the dollar values at the 50th iteration are about 2 cents different from those by GLM. However, using our algorithm, the GMBM model converges completely in 5 iterations with solutions identical to GLM results.

Appendix 2: Numerical Results of GMBM

Table 1: The Age and Vehicle-use Relativities for Selected GMBMs with k=0.5

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
0	-2.5	1.336	1.231	1.228	1.159	0.885	1.026	1.035	1.000	1.744	1.288	1.093	1.000
0	**-2**	**1.348**	**1.227**	**1.219**	**1.156**	**0.884**	**1.023**	**1.032**	**1.000**	**1.749**	**1.282**	**1.090**	**1.000**
0	-1.5	1.365	1.223	1.210	1.152	0.882	1.019	1.029	1.000	1.755	1.275	1.089	1.000
0	-1	1.386	1.216	1.199	1.148	0.879	1.017	1.026	1.000	1.763	1.269	1.088	1.000
0	-0.5	1.414	1.209	1.188	1.144	0.874	1.014	1.023	1.000	1.773	1.263	1.087	1.000
0	0	1.448	1.201	1.175	1.139	0.869	1.012	1.020	1.000	1.788	1.256	1.086	1.000
0	0.5	1.490	1.191	1.162	1.134	0.862	1.011	1.017	1.000	1.809	1.248	1.086	1.000
0	1	1.543	1.180	1.147	1.129	0.854	1.010	1.014	1.000	1.839	1.238	1.086	1.000
0	1.5	1.611	1.167	1.130	1.122	0.844	1.010	1.011	1.000	1.886	1.224	1.086	1.000
0	2	1.700	1.151	1.110	1.115	0.831	1.011	1.008	1.000	1.959	1.206	1.086	1.000
0.5	-1.5	1.313	1.270	1.217	1.156	0.912	1.014	1.028	1.000	1.665	1.280	1.063	1.000
0.5	**-1**	**1.320**	**1.263**	**1.209**	**1.154**	**0.909**	**1.012**	**1.026**	**1.000**	**1.665**	**1.277**	**1.062**	**1.000**
0.5	-0.5	1.328	1.256	1.200	1.151	0.905	1.010	1.024	1.000	1.665	1.274	1.062	1.000
0.5	0	1.340	1.247	1.191	1.148	0.900	1.009	1.021	1.000	1.667	1.272	1.062	1.000
0.5	0.5	1.356	1.237	1.181	1.144	0.894	1.007	1.019	1.000	1.669	1.269	1.062	1.000
0.5	1	1.376	1.226	1.170	1.140	0.887	1.006	1.017	1.000	1.673	1.266	1.062	1.000
0.5	1.5	1.401	1.215	1.158	1.136	0.879	1.005	1.014	1.000	1.680	1.263	1.062	1.000
0.5	2	1.433	1.202	1.145	1.131	0.870	1.004	1.012	1.000	1.691	1.260	1.063	1.000
1	-1	1.291	1.313	1.220	1.160	0.938	1.010	1.026	1.000	1.645	1.267	1.043	1.000
1	**-0.5**	**1.291**	**1.305**	**1.213**	**1.158**	**0.934**	**1.008**	**1.024**	**1.000**	**1.643**	**1.266**	**1.043**	**1.000**
1	0	1.291	1.297	1.206	1.156	0.929	1.007	1.022	1.000	1.642	1.265	1.043	1.000
1	0.5	1.294	1.287	1.198	1.154	0.924	1.005	1.021	1.000	1.640	1.264	1.043	1.000
1	1	1.298	1.276	1.190	1.152	0.918	1.004	1.019	1.000	1.639	1.263	1.043	1.000
1	1.5	1.305	1.265	1.181	1.149	0.911	1.004	1.017	1.000	1.638	1.262	1.042	1.000
1	2	1.314	1.253	1.171	1.146	0.904	1.003	1.015	1.000	1.638	1.261	1.042	1.000

Table 1: The Age and Vehicle-use Relativities for Selected GMBMs with k=0.5, continued

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1.5	0	1.281	1.343	1.218	1.162	0.953	1.006	1.024	1.000	1.643	1.253	1.030	1.000
1.5	0.5	1.277	1.334	1.211	1.161	0.948	1.005	1.022	1.000	1.641	1.252	1.030	1.000
1.5	**1**	**1.274**	**1.324**	**1.205**	**1.159**	**0.943**	**1.003**	**1.021**	**1.000**	**1.639**	**1.252**	**1.030**	**1.000**
1.5	1.5	1.272	1.313	1.197	1.157	0.937	1.002	1.019	1.000	1.638	1.252	1.030	1.000
1.5	2	1.271	1.301	1.190	1.155	0.930	1.002	1.018	1.000	1.636	1.252	1.029	1.000
2	0	1.288	1.385	1.228	1.167	0.972	1.007	1.025	1.000	1.651	1.239	1.022	1.000
2	0.5	1.282	1.376	1.223	1.166	0.968	1.006	1.024	1.000	1.650	1.240	1.022	1.000
2	1	1.277	1.367	1.217	1.165	0.963	1.004	1.023	1.000	1.648	1.240	1.022	1.000
2	1.5	1.272	1.357	1.211	1.163	0.958	1.003	1.022	1.000	1.646	1.240	1.022	1.000
2	2	1.266	1.346	1.205	1.162	0.952	1.002	1.020	1.000	1.644	1.241	1.021	1.000
2	2.5	1.262	1.335	1.198	1.160	0.945	1.001	1.019	1.000	1.643	1.241	1.021	1.000
2	**3**	**1.258**	**1.322**	**1.190**	**1.157**	**0.938**	**1.000**	**1.018**	**1.000**	**1.641**	**1.241**	**1.020**	**1.000**
2	3.5	1.255	1.308	1.182	1.155	0.930	0.999	1.016	1.000	1.640	1.241	1.020	1.000

Table 2: The Age and Vehicle-use Relativities for Selected GMBMs with k=1

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
0	-2.5	1.353	1.235	1.229	1.159	0.887	1.027	1.036	1.000	1.757	1.290	1.093	1.000
0	**-2**	**1.369**	**1.231**	**1.221**	**1.156**	**0.886**	**1.023**	**1.032**	**1.000**	**1.762**	**1.284**	**1.090**	**1.000**
0	-1.5	1.389	1.226	1.212	1.153	0.884	1.020	1.029	1.000	1.768	1.278	1.089	1.000
0	-1	1.414	1.220	1.201	1.149	0.881	1.017	1.026	1.000	1.776	1.272	1.088	1.000
0	-0.5	1.445	1.213	1.190	1.144	0.877	1.014	1.023	1.000	1.787	1.267	1.087	1.000
0	0	1.483	1.204	1.178	1.140	0.872	1.012	1.020	1.000	1.801	1.260	1.087	1.000
0	0.5	1.528	1.195	1.165	1.135	0.865	1.010	1.017	1.000	1.823	1.253	1.087	1.000
0	1	1.582	1.184	1.150	1.129	0.858	1.009	1.014	1.000	1.854	1.244	1.087	1.000
0	1.5	1.649	1.171	1.134	1.123	0.848	1.009	1.011	1.000	1.899	1.233	1.087	1.000
0	2	1.731	1.156	1.116	1.116	0.836	1.010	1.008	1.000	1.967	1.217	1.087	1.000
0.5	-1.5	1.330	1.275	1.218	1.156	0.914	1.014	1.028	1.000	1.670	1.280	1.062	1.000
0.5	**-1**	**1.340**	**1.268**	**1.210**	**1.153**	**0.911**	**1.012**	**1.026**	**1.000**	**1.670**	**1.277**	**1.062**	**1.000**
0.5	-0.5	1.352	1.260	1.201	1.150	0.906	1.010	1.023	1.000	1.671	1.274	1.061	1.000
0.5	0	1.368	1.250	1.191	1.147	0.901	1.009	1.021	1.000	1.673	1.272	1.061	1.000
0.5	0.5	1.388	1.240	1.181	1.144	0.895	1.007	1.019	1.000	1.676	1.270	1.062	1.000
0.5	1	1.412	1.230	1.170	1.140	0.889	1.006	1.016	1.000	1.681	1.267	1.062	1.000
0.5	1.5	1.442	1.218	1.158	1.135	0.881	1.005	1.014	1.000	1.689	1.265	1.063	1.000
0.5	2	1.479	1.205	1.146	1.130	0.872	1.004	1.011	1.000	1.702	1.262	1.064	1.000
1	-1	1.303	1.318	1.220	1.159	0.939	1.010	1.026	1.000	1.647	1.266	1.042	1.000
1	**-0.5**	**1.304**	**1.310**	**1.213**	**1.158**	**0.935**	**1.008**	**1.024**	**1.000**	**1.646**	**1.265**	**1.042**	**1.000**
1	0	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000
1	0.5	1.312	1.291	1.198	1.154	0.925	1.006	1.020	1.000	1.643	1.263	1.042	1.000
1	1	1.319	1.280	1.190	1.151	0.919	1.005	1.019	1.000	1.642	1.262	1.042	1.000
1	1.5	1.329	1.269	1.181	1.148	0.912	1.004	1.017	1.000	1.641	1.261	1.042	1.000
1	2	1.343	1.256	1.171	1.145	0.905	1.003	1.015	1.000	1.641	1.260	1.042	1.000
1.5	0	1.290	1.348	1.218	1.162	0.955	1.006	1.023	1.000	1.645	1.251	1.029	1.000
1.5	0.5	1.287	1.339	1.212	1.160	0.950	1.005	1.022	1.000	1.643	1.251	1.029	1.000
1.5	**1**	**1.286**	**1.328**	**1.205**	**1.159**	**0.944**	**1.004**	**1.021**	**1.000**	**1.641**	**1.251**	**1.029**	**1.000**

Table 2: The Age and Vehicle-use Relativities for Selected GMBMs with k=1, continued

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1.5	1.5	1.286	1.317	1.198	1.157	0.938	1.003	1.019	1.000	1.639	1.251	1.029	1.000
1.5	2	1.287	1.305	1.190	1.154	0.931	1.002	1.018	1.000	1.638	1.251	1.029	1.000
2	0	1.294	1.389	1.228	1.167	0.973	1.007	1.025	1.000	1.653	1.238	1.021	1.000
2	0.5	1.289	1.380	1.223	1.166	0.969	1.006	1.024	1.000	1.652	1.238	1.021	1.000
2	1	1.284	1.371	1.217	1.165	0.964	1.004	1.023	1.000	1.650	1.239	1.021	1.000
2	1.5	1.280	1.361	1.211	1.163	0.959	1.003	1.021	1.000	1.648	1.239	1.020	1.000
2	2	1.276	1.351	1.205	1.161	0.953	1.002	1.020	1.000	1.646	1.239	1.020	1.000
2	2.5	1.272	1.339	1.198	1.159	0.946	1.001	1.019	1.000	1.644	1.240	1.020	1.000
2	3	1.270	1.326	1.190	1.157	0.939	1.000	1.017	1.000	1.643	1.240	1.020	1.000
2	**3.5**	**1.268**	**1.312**	**1.182**	**1.154**	**0.931**	**0.999**	**1.016**	**1.000**	**1.641**	**1.240**	**1.019**	**1.000**
2	4	1.269	1.298	1.173	1.151	0.922	0.999	1.014	1.000	1.640	1.240	1.018	1.000

Table 3: The Age and Vehicle-use Relativities for Selected GMBMs with k=1.5

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
0	-2	1.391	1.235	1.223	1.156	0.888	1.024	1.033	1.000	1.774	1.286	1.090	1.000
0	**-1.5**	**1.415**	**1.230**	**1.214**	**1.153**	**0.886**	**1.020**	**1.030**	**1.000**	**1.780**	**1.281**	**1.089**	**1.000**
0	-1	1.444	1.224	1.203	1.149	0.884	1.017	1.026	1.000	1.788	1.275	1.088	1.000
0	-0.5	1.478	1.217	1.192	1.145	0.880	1.014	1.023	1.000	1.798	1.270	1.087	1.000
0	0	1.519	1.208	1.181	1.140	0.875	1.012	1.020	1.000	1.813	1.264	1.087	1.000
0	0.5	1.566	1.199	1.168	1.135	0.869	1.010	1.017	1.000	1.834	1.258	1.087	1.000
0	1	1.621	1.188	1.154	1.130	0.861	1.009	1.014	1.000	1.865	1.250	1.088	1.000
0	1.5	1.685	1.176	1.138	1.124	0.852	1.009	1.011	1.000	1.908	1.241	1.088	1.000
0	2	1.762	1.162	1.121	1.118	0.842	1.009	1.008	1.000	1.972	1.228	1.089	1.000
0.5	-1.5	1.350	1.279	1.218	1.156	0.916	1.015	1.028	1.000	1.675	1.279	1.061	1.000
0.5	**-1**	**1.362**	**1.272**	**1.210**	**1.153**	**0.912**	**1.012**	**1.025**	**1.000**	**1.675**	**1.277**	**1.061**	**1.000**
0.5	-0.5	1.378	1.264	1.202	1.150	0.908	1.010	1.023	1.000	1.677	1.274	1.061	1.000
0.5	0	1.398	1.254	1.192	1.147	0.903	1.009	1.021	1.000	1.679	1.272	1.061	1.000
0.5	0.5	1.422	1.244	1.182	1.143	0.897	1.007	1.018	1.000	1.682	1.270	1.061	1.000
0.5	1	1.451	1.233	1.171	1.139	0.890	1.006	1.016	1.000	1.688	1.268	1.062	1.000
0.5	1.5	1.485	1.221	1.159	1.134	0.883	1.005	1.013	1.000	1.698	1.266	1.063	1.000
0.5	2	1.526	1.208	1.146	1.129	0.874	1.004	1.011	1.000	1.712	1.264	1.065	1.000
1	-1	1.315	1.323	1.220	1.159	0.941	1.010	1.025	1.000	1.650	1.264	1.041	1.000
1	**-0.5**	**1.319**	**1.315**	**1.214**	**1.157**	**0.937**	**1.008**	**1.024**	**1.000**	**1.648**	**1.264**	**1.041**	**1.000**
1	0	1.325	1.306	1.206	1.155	0.932	1.007	1.022	1.000	1.647	1.263	1.041	1.000
1	0.5	1.333	1.295	1.198	1.153	0.927	1.006	1.020	1.000	1.645	1.262	1.041	1.000
1	1	1.344	1.284	1.190	1.150	0.921	1.005	1.018	1.000	1.644	1.261	1.041	1.000
1	1.5	1.357	1.273	1.181	1.148	0.914	1.004	1.016	1.000	1.644	1.261	1.041	1.000
1	2	1.375	1.260	1.171	1.144	0.906	1.003	1.015	1.000	1.644	1.260	1.042	1.000
1.5	0	1.299	1.353	1.218	1.162	0.956	1.006	1.023	1.000	1.647	1.250	1.028	1.000
1.5	0.5	1.299	1.343	1.212	1.160	0.951	1.005	1.022	1.000	1.645	1.250	1.028	1.000
1.5	**1**	**1.299**	**1.333**	**1.205**	**1.158**	**0.945**	**1.004**	**1.020**	**1.000**	**1.643**	**1.250**	**1.028**	**1.000**

Table 3: The Age and Vehicle-use Relativities for Selected GMBMs with k=1.5, continued

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1.5	1.5	1.301	1.322	1.198	1.156	0.939	1.003	1.019	1.000	1.641	1.250	1.028	1.000
1.5	2	1.305	1.310	1.190	1.154	0.933	1.002	1.017	1.000	1.639	1.250	1.028	1.000
2	0	1.300	1.393	1.228	1.167	0.974	1.007	1.025	1.000	1.655	1.236	1.019	1.000
2	0.5	1.295	1.384	1.223	1.166	0.970	1.005	1.024	1.000	1.654	1.237	1.019	1.000
2	1	1.292	1.375	1.217	1.164	0.965	1.004	1.022	1.000	1.652	1.237	1.019	1.000
2	1.5	1.288	1.366	1.211	1.163	0.960	1.003	1.021	1.000	1.650	1.238	1.019	1.000
2	2	1.286	1.355	1.205	1.161	0.954	1.002	1.020	1.000	1.648	1.238	1.019	1.000
2	2.5	1.284	1.343	1.198	1.159	0.947	1.001	1.018	1.000	1.646	1.239	1.019	1.000
2	3	1.283	1.330	1.190	1.156	0.940	1.000	1.017	1.000	1.644	1.239	1.019	1.000
2	**3.5**	**1.284**	**1.316**	**1.182**	**1.153**	**0.932**	**0.999**	**1.016**	**1.000**	**1.642**	**1.239**	**1.018**	**1.000**
2	4	1.286	1.302	1.174	1.150	0.924	0.999	1.014	1.000	1.641	1.239	1.018	1.000

Table 4: The Age and Vehicle-use Relativities for Selected GMBMs with k=2

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DriveLong	DriveShort	Pleasure
0	-2	1.414	1.240	1.225	1.157	0.891	1.024	1.033	1.000	1.786	1.289	1.090	1.000
0	**-1.5**	**1.442**	**1.235**	**1.216**	**1.153**	**0.889**	**1.021**	**1.030**	**1.000**	**1.791**	**1.283**	**1.089**	**1.000**
0	-1	1.475	1.229	1.206	1.149	0.886	1.017	1.027	1.000	1.798	1.278	1.088	1.000
0	-0.5	1.513	1.221	1.195	1.145	0.882	1.015	1.023	1.000	1.808	1.274	1.087	1.000
0	0	1.556	1.213	1.183	1.141	0.878	1.012	1.020	1.000	1.822	1.268	1.087	1.000
0	0.5	1.605	1.203	1.171	1.136	0.872	1.010	1.017	1.000	1.843	1.263	1.088	1.000
0	1	1.659	1.193	1.157	1.131	0.865	1.009	1.014	1.000	1.873	1.256	1.088	1.000
0	1.5	1.721	1.181	1.143	1.125	0.857	1.008	1.011	1.000	1.914	1.248	1.089	1.000
0	2	1.791	1.168	1.127	1.119	0.847	1.008	1.008	1.000	1.973	1.238	1.090	1.000
0.5	-1.5	1.371	1.284	1.219	1.155	0.918	1.015	1.028	1.000	1.680	1.279	1.060	1.000
0.5	**-1**	**1.387**	**1.277**	**1.211**	**1.153**	**0.914**	**1.013**	**1.025**	**1.000**	**1.681**	**1.276**	**1.060**	**1.000**
0.5	-0.5	1.407	1.268	1.202	1.150	0.910	1.010	1.023	1.000	1.682	1.274	1.060	1.000
0.5	0	1.431	1.258	1.193	1.146	0.905	1.009	1.021	1.000	1.684	1.273	1.060	1.000
0.5	0.5	1.459	1.248	1.183	1.143	0.899	1.007	1.018	1.000	1.689	1.271	1.061	1.000
0.5	1	1.492	1.237	1.172	1.139	0.892	1.006	1.016	1.000	1.695	1.269	1.062	1.000
0.5	1.5	1.530	1.224	1.160	1.134	0.884	1.004	1.013	1.000	1.705	1.268	1.064	1.000
0.5	2	1.573	1.211	1.147	1.129	0.876	1.003	1.011	1.000	1.721	1.266	1.065	1.000
1	-1	1.330	1.328	1.220	1.159	0.943	1.010	1.025	1.000	1.652	1.263	1.039	1.000
1	-0.5	1.336	1.320	1.214	1.157	0.939	1.008	1.024	1.000	1.651	1.262	1.039	1.000
1	0	1.345	1.310	1.206	1.155	0.934	1.007	1.022	1.000	1.649	1.262	1.040	1.000
1	0.5	1.356	1.300	1.199	1.153	0.928	1.006	1.020	1.000	1.648	1.261	1.040	1.000
1	1	1.371	1.289	1.190	1.150	0.922	1.005	1.018	1.000	1.647	1.261	1.040	1.000
1	1.5	1.389	1.277	1.181	1.147	0.915	1.004	1.016	1.000	1.647	1.260	1.041	1.000
1	2	1.410	1.264	1.172	1.144	0.908	1.003	1.014	1.000	1.648	1.260	1.041	1.000
1.5	0	1.310	1.357	1.218	1.161	0.957	1.006	1.023	1.000	1.649	1.248	1.026	1.000
1.5	0.5	1.311	1.348	1.212	1.160	0.952	1.005	1.022	1.000	1.647	1.248	1.027	1.000
1.5	**1**	**1.314**	**1.337**	**1.205**	**1.158**	**0.947**	**1.004**	**1.020**	**1.000**	**1.645**	**1.249**	**1.027**	**1.000**
1.5	1.5	1.319	1.326	1.198	1.156	0.941	1.003	1.019	1.000	1.643	1.249	1.027	1.000

Table 4: The Age and Vehicle-use Relativities for Selected GMBMs with k=2, continued

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DriveLong	DriveShort	Pleasure
1.5	2	1.326	1.314	1.190	1.153	0.934	1.002	1.017	1.000	1.641	1.249	1.027	1.000
2	0	1.306	1.396	1.228	1.167	0.975	1.007	1.024	1.000	1.657	1.235	1.018	1.000
2	0.5	1.303	1.388	1.223	1.165	0.971	1.005	1.023	1.000	1.655	1.236	1.018	1.000
2	1	1.300	1.379	1.217	1.164	0.966	1.004	1.022	1.000	1.654	1.236	1.018	1.000
2	1.5	1.298	1.369	1.212	1.162	0.961	1.003	1.021	1.000	1.652	1.237	1.018	1.000
2	2	1.297	1.359	1.205	1.160	0.955	1.001	1.020	1.000	1.650	1.237	1.018	1.000
2	2.5	1.297	1.347	1.198	1.158	0.948	1.001	1.018	1.000	1.648	1.238	1.018	1.000
2	3	1.298	1.334	1.191	1.156	0.941	1.000	1.017	1.000	1.646	1.238	1.018	1.000
2	**3.5**	**1.301**	**1.321**	**1.183**	**1.153**	**0.934**	**0.999**	**1.015**	**1.000**	**1.644**	**1.238**	**1.018**	**1.000**
2	4	1.307	1.306	1.174	1.150	0.925	0.999	1.014	1.000	1.642	1.238	1.017	1.000

Table 5: The Age and Vehicle-use Relativities for Selected GMBMs with k=2.5

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DriveLong	DriveShort	Pleasure
0	-2	1.439	1.244	1.227	1.157	0.893	1.025	1.033	1.000	1.797	1.290	1.090	1.000
0	**-1.5**	**1.471**	**1.239**	**1.218**	**1.154**	**0.891**	**1.021**	**1.030**	**1.000**	**1.801**	**1.285**	**1.088**	**1.000**
0	-1	1.507	1.233	1.208	1.150	0.888	1.018	1.027	1.000	1.807	1.281	1.087	1.000
0	-0.5	1.548	1.226	1.197	1.146	0.885	1.015	1.024	1.000	1.816	1.276	1.087	1.000
0	0	1.594	1.217	1.186	1.141	0.880	1.012	1.020	1.000	1.829	1.272	1.087	1.000
0	0.5	1.643	1.208	1.174	1.137	0.875	1.010	1.017	1.000	1.849	1.267	1.088	1.000
0	1	1.697	1.198	1.161	1.132	0.869	1.009	1.014	1.000	1.877	1.261	1.089	1.000
0	1.5	1.755	1.186	1.147	1.126	0.861	1.008	1.011	1.000	1.917	1.255	1.089	1.000
0	2	1.819	1.174	1.132	1.121	0.852	1.008	1.008	1.000	1.972	1.246	1.090	1.000
0.5	-2	1.379	1.296	1.227	1.157	0.923	1.017	1.030	1.000	1.684	1.280	1.059	1.000
0.5	**-1.5**	**1.395**	**1.289**	**1.219**	**1.155**	**0.920**	**1.015**	**1.028**	**1.000**	**1.685**	**1.278**	**1.059**	**1.000**
0.5	-1	1.415	1.281	1.212	1.152	0.916	1.013	1.025	1.000	1.686	1.276	1.059	1.000
0.5	-0.5	1.439	1.273	1.203	1.149	0.912	1.011	1.023	1.000	1.687	1.274	1.059	1.000
0.5	0	1.467	1.263	1.194	1.146	0.907	1.009	1.020	1.000	1.690	1.273	1.060	1.000
0.5	0.5	1.499	1.252	1.184	1.142	0.901	1.007	1.018	1.000	1.694	1.271	1.061	1.000
0.5	1	1.535	1.241	1.173	1.138	0.894	1.006	1.015	1.000	1.701	1.270	1.062	1.000
0.5	1.5	1.576	1.228	1.161	1.133	0.886	1.004	1.013	1.000	1.712	1.269	1.064	1.000
0.5	2	1.621	1.215	1.149	1.128	0.878	1.003	1.010	1.000	1.728	1.268	1.066	1.000
1	-1	1.346	1.333	1.221	1.159	0.944	1.010	1.025	1.000	1.655	1.262	1.038	1.000
1	**-0.5**	**1.355**	**1.324**	**1.214**	**1.157**	**0.940**	**1.008**	**1.023**	**1.000**	**1.653**	**1.261**	**1.038**	**1.000**
1	0	1.367	1.315	1.207	1.155	0.935	1.007	1.022	1.000	1.652	1.261	1.038	1.000
1	0.5	1.382	1.304	1.199	1.152	0.930	1.006	1.020	1.000	1.651	1.260	1.039	1.000
1	1	1.401	1.293	1.191	1.149	0.924	1.005	1.018	1.000	1.650	1.260	1.040	1.000
1	1.5	1.423	1.281	1.181	1.146	0.917	1.004	1.016	1.000	1.650	1.260	1.040	1.000
1	2	1.449	1.268	1.172	1.143	0.909	1.003	1.014	1.000	1.652	1.259	1.041	1.000
1.5	0	1.322	1.361	1.218	1.161	0.958	1.006	1.023	1.000	1.651	1.247	1.025	1.000
1.5	0.5	1.325	1.352	1.212	1.159	0.954	1.005	1.021	1.000	1.649	1.247	1.025	1.000
1.5	1	1.331	1.342	1.205	1.157	0.948	1.003	1.020	1.000	1.647	1.247	1.026	1.000

Table 5: The Age and Vehicle-use Relativities for Selected GMBMs with k=2.5, continued

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DriveLong	DriveShort	Pleasure
1.5	**1.5**	**1.338**	**1.330**	**1.198**	**1.155**	**0.942**	**1.003**	**1.018**	**1.000**	**1.645**	**1.248**	**1.026**	**1.000**
1.5	2	1.348	1.318	1.190	1.153	0.935	1.002	1.017	1.000	1.643	1.248	1.026	1.000
2	0	1.312	1.400	1.228	1.166	0.976	1.007	1.024	1.000	1.659	1.234	1.017	1.000
2	0.5	1.310	1.392	1.223	1.165	0.972	1.005	1.023	1.000	1.657	1.235	1.017	1.000
2	1	1.309	1.383	1.217	1.164	0.967	1.004	1.022	1.000	1.655	1.235	1.017	1.000
2	1.5	1.308	1.373	1.212	1.162	0.962	1.002	1.021	1.000	1.653	1.236	1.017	1.000
2	2	1.309	1.363	1.205	1.160	0.956	1.001	1.019	1.000	1.651	1.236	1.017	1.000
2	2.5	1.311	1.351	1.198	1.158	0.950	1.000	1.018	1.000	1.649	1.237	1.017	1.000
2	3	1.315	1.338	1.191	1.155	0.942	1.000	1.016	1.000	1.647	1.237	1.017	1.000
2	**3.5**	**1.321**	**1.325**	**1.183**	**1.152**	**0.935**	**0.999**	**1.015**	**1.000**	**1.645**	**1.237**	**1.017**	**1.000**
2	4	1.329	1.310	1.174	1.149	0.927	0.999	1.013	1.000	1.644	1.238	1.017	1.000

Table 6: The Age and Vehicle-use Relativities for Selected GMBMs with k=3

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DriveLong	DriveShort	Pleasure
0	-2	1.466	1.249	1.228	1.158	0.895	1.025	1.034	1.000	1.806	1.292	1.089	1.000
0	**-1.5**	**1.501**	**1.244**	**1.220**	**1.154**	**0.893**	**1.022**	**1.030**	**1.000**	**1.808**	**1.287**	**1.088**	**1.000**
0	-1	1.541	1.238	1.210	1.150	0.891	1.018	1.027	1.000	1.813	1.283	1.087	1.000
0	-0.5	1.585	1.230	1.200	1.146	0.887	1.015	1.024	1.000	1.820	1.279	1.087	1.000
0	0	1.631	1.222	1.189	1.142	0.883	1.013	1.020	1.000	1.833	1.275	1.087	1.000
0	0.5	1.681	1.212	1.177	1.137	0.878	1.010	1.017	1.000	1.852	1.271	1.088	1.000
0	1	1.733	1.202	1.165	1.132	0.872	1.009	1.014	1.000	1.879	1.266	1.089	1.000
0	1.5	1.787	1.191	1.152	1.127	0.865	1.007	1.011	1.000	1.917	1.260	1.090	1.000
0	2	1.845	1.179	1.137	1.122	0.857	1.007	1.008	1.000	1.968	1.253	1.091	1.000
0.5	-2	1.401	1.301	1.227	1.157	0.925	1.017	1.030	1.000	1.689	1.280	1.058	1.000
0.5	**-1.5**	**1.421**	**1.294**	**1.220**	**1.155**	**0.922**	**1.015**	**1.027**	**1.000**	**1.689**	**1.278**	**1.057**	**1.000**
0.5	-1	1.445	1.286	1.212	1.152	0.918	1.013	1.025	1.000	1.690	1.276	1.058	1.000
0.5	-0.5	1.472	1.277	1.204	1.149	0.914	1.011	1.023	1.000	1.691	1.274	1.058	1.000
0.5	0	1.504	1.267	1.194	1.146	0.909	1.009	1.020	1.000	1.694	1.273	1.059	1.000
0.5	0.5	1.540	1.256	1.184	1.142	0.903	1.007	1.018	1.000	1.699	1.272	1.060	1.000
0.5	1	1.578	1.245	1.174	1.138	0.896	1.006	1.015	1.000	1.706	1.271	1.062	1.000
0.5	1.5	1.621	1.232	1.162	1.133	0.889	1.004	1.013	1.000	1.717	1.270	1.064	1.000
0.5	2	1.667	1.219	1.150	1.128	0.880	1.003	1.010	1.000	1.734	1.270	1.066	1.000
1	-1	1.364	1.338	1.221	1.158	0.946	1.010	1.025	1.000	1.657	1.260	1.036	1.000
1	**-0.5**	**1.377**	**1.329**	**1.214**	**1.156**	**0.942**	**1.008**	**1.023**	**1.000**	**1.656**	**1.260**	**1.037**	**1.000**
1	0	1.392	1.319	1.207	1.154	0.937	1.007	1.021	1.000	1.655	1.260	1.037	1.000
1	0.5	1.411	1.309	1.199	1.152	0.931	1.006	1.019	1.000	1.654	1.259	1.038	1.000
1	1	1.434	1.297	1.191	1.149	0.925	1.005	1.017	1.000	1.653	1.259	1.039	1.000
1	1.5	1.460	1.285	1.182	1.146	0.918	1.004	1.015	1.000	1.654	1.259	1.040	1.000
1	2	1.490	1.272	1.172	1.142	0.911	1.003	1.013	1.000	1.655	1.259	1.041	1.000
1.5	0	1.335	1.366	1.218	1.161	0.960	1.006	1.023	1.000	1.652	1.246	1.024	1.000
1.5	0.5	1.341	1.356	1.212	1.159	0.955	1.005	1.021	1.000	1.650	1.246	1.024	1.000
1.5	1	1.349	1.346	1.205	1.157	0.949	1.003	1.020	1.000	1.648	1.246	1.024	1.000

Table 6: The Age and Vehicle-use Relativities for Selected GMBMs with k=3, continued

p	q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DriveLong	DriveShort	Pleasure
1.5	**1.5**	**1.360**	**1.334**	**1.198**	**1.155**	**0.943**	**1.002**	**1.018**	**1.000**	**1.647**	**1.247**	**1.025**	**1.000**
1.5	2	1.374	1.322	1.190	1.152	0.936	1.002	1.016	1.000	1.645	1.247	1.025	1.000
2	0	1.319	1.404	1.228	1.166	0.977	1.007	1.024	1.000	1.661	1.233	1.016	1.000
2	0.5	1.318	1.396	1.223	1.165	0.973	1.005	1.023	1.000	1.659	1.233	1.016	1.000
2	1	1.318	1.387	1.217	1.163	0.968	1.004	1.022	1.000	1.657	1.234	1.016	1.000
2	1.5	1.320	1.377	1.212	1.162	0.963	1.002	1.020	1.000	1.655	1.235	1.016	1.000
2	2	1.322	1.366	1.205	1.160	0.957	1.001	1.019	1.000	1.653	1.235	1.016	1.000
2	2.5	1.327	1.355	1.198	1.157	0.951	1.000	1.018	1.000	1.651	1.236	1.016	1.000
2	3	1.334	1.342	1.191	1.155	0.944	1.000	1.016	1.000	1.649	1.236	1.016	1.000
2	**3.5**	**1.343**	**1.329**	**1.183**	**1.152**	**0.936**	**0.999**	**1.014**	**1.000**	**1.647**	**1.237**	**1.016**	**1.000**
2	4	1.354	1.314	1.175	1.149	0.928	0.999	1.013	1.000	1.645	1.237	1.016	1.000

Table 7: The Age and Vehicle-use Relativities for Selected GLMs

Distribution	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
Inverse Gaussian	1.303	1.318	1.220	1.159	0.939	1.010	1.026	1.000	1.647	1.266	1.042	1.000
Gamma	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000
Poisson	1.319	1.280	1.190	1.151	0.919	1.005	1.019	1.000	1.642	1.262	1.042	1.000
Normal	1.343	1.256	1.171	1.145	0.905	1.003	1.015	1.000	1.641	1.260	1.042	1.000

Table 8: The Age and Vehicle-use Relativities for Selected GMBMs with k=1 and p=1

q	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
-1	1.303	1.318	1.220	1.159	0.939	1.010	1.026	1.000	1.647	1.266	1.042	1.000
-0.5	1.304	1.310	1.213	1.158	0.935	1.008	1.024	1.000	1.646	1.265	1.042	1.000
0	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000
0.5	1.312	1.291	1.198	1.154	0.925	1.006	1.020	1.000	1.643	1.263	1.042	1.000
1	1.319	1.280	1.190	1.151	0.919	1.005	1.019	1.000	1.642	1.262	1.042	1.000
1.5	1.329	1.269	1.181	1.148	0.912	1.004	1.017	1.000	1.641	1.261	1.042	1.000
2	1.343	1.256	1.171	1.145	0.905	1.003	1.015	1.000	1.641	1.260	1.042	1.000

Table 9: The Model Performance of GMBM with K=0.5

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2.5	13.747	5.63%	1.262	4.1651
0	**-2**	**13.545**	**5.56%**	**1.255**	**4.1225**
0	-1.5	13.588	5.58%	1.260	4.1384
0	-1	13.770	5.65%	1.283	4.2024
0	-0.5	14.050	5.77%	1.328	4.3198
0	0	14.618	6.01%	1.411	4.5418
0	0.5	15.337	6.32%	1.559	4.8894
0	1	16.467	6.80%	1.832	5.4920
0	1.5	18.211	7.58%	2.382	6.5860
0	2	21.788	9.33%	3.642	8.9078
0.5	-1.5	11.593	4.67%	1.056	3.4992
0.5	**-1**	**11.584**	**4.69%**	**1.052**	**3.4912**
0.5	-0.5	11.735	4.76%	1.052	3.5131
0.5	0	11.972	4.88%	1.056	3.5559
0.5	0.5	12.261	5.01%	1.068	3.6180
0.5	1	12.594	5.17%	1.089	3.7036
0.5	1.5	12.974	5.35%	1.126	3.8221
0.5	2	13.410	5.57%	1.187	3.9900
1	-1	10.696	4.18%	1.046	3.3453
1	**-0.5**	**10.739**	**4.22%**	**1.040**	**3.3420**
1	0	10.851	4.28%	1.035	3.3506
1	0.5	11.018	4.37%	1.031	3.3699
1	1	11.208	4.47%	1.029	3.3967
1	1.5	11.422	4.59%	1.032	3.4330
1	2	11.658	4.72%	1.040	3.4818
1.5	0	10.376	3.96%	1.077	3.3435
1.5	0.5	10.452	4.01%	1.067	3.3391
<u>1.5</u>	**<u>1</u>**	**<u>10.531</u>**	**<u>4.06%</u>**	**<u>1.057</u>**	**<u>3.3359</u>**

1.5	1.5	10.657	4.14%	1.048	3.3419
1.5	2	10.820	4.23%	1.042	3.3572
2	0	10.310	3.84%	1.155	3.4515
2	0.5	10.374	3.88%	1.140	3.4389
2	1	10.443	3.93%	1.124	3.4262
2	1.5	10.518	3.98%	1.108	3.4143
2	2	10.600	4.04%	1.093	3.4040
2	2.5	10.688	4.10%	1.080	3.3969
2	**3**	**10.781**	**4.16%**	**1.069**	**3.3943**
2	3.5	10.877	4.23%	1.062	3.3982

Table 10: The Model Performance of GMBM with K=1

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2.5	13.873	5.65%	1.276	4.2068
0	**-2**	**13.631**	**5.56%**	**1.268**	**4.1575**
0	-1.5	13.614	5.56%	1.274	4.1642
0	-1	13.753	5.61%	1.296	4.2221
0	-0.5	14.018	5.72%	1.343	4.3381
0	0	14.588	5.96%	1.426	4.5612
0	0.5	15.344	6.27%	1.573	4.9128
0	1	16.480	6.74%	1.838	5.5038
0	1.5	18.011	7.43%	2.353	6.5095
0	2	21.263	9.00%	3.460	8.5778
0.5	-1.5	11.563	4.64%	1.049	3.4825
0.5	**-1**	**11.554**	**4.66%**	**1.045**	**3.4746**
0.5	-0.5	11.722	4.74%	1.045	3.4998
0.5	0	11.977	4.86%	1.050	3.5470
0.5	0.5	12.274	5.00%	1.064	3.6134
0.5	1	12.613	5.16%	1.089	3.7054
0.5	1.5	12.998	5.34%	1.130	3.8331
0.5	2	13.434	5.56%	1.199	4.0139

1	-1	10.669	4.15%	1.043	3.3358
1	**-0.5**	**10.716**	**4.19%**	**1.036**	**3.3313**
1	0	10.826	4.26%	1.029	3.3376
1	0.5	10.996	4.35%	1.024	3.3556
1	1	11.190	4.45%	1.022	3.3815
1	1.5	11.408	4.57%	1.024	3.4178
1	2	11.664	4.70%	1.032	3.4696
1.5	0	10.350	3.94%	1.079	3.3423
1.5	0.5	10.424	3.99%	1.067	3.3353
1.5	**1**	**10.503**	**4.04%**	**1.055**	**3.3295**
1.5	1.5	10.637	4.12%	1.045	3.3340
1.5	2	10.804	4.21%	1.037	3.3471
2	0	10.290	3.82%	1.164	3.4614
2	0.5	10.353	3.86%	1.148	3.4471
2	1	10.422	3.91%	1.130	3.4323
2	1.5	10.495	3.96%	1.113	3.4179
2	2	10.577	4.01%	1.096	3.4051
2	2.5	10.663	4.07%	1.081	3.3947
2	3	10.753	4.14%	1.068	3.3883
2	**3.5**	**10.846**	**4.21%**	**1.058**	**3.3881**
2	4	10.983	4.30%	1.054	3.4028

Table 11: The Model Performance of GMBM with K=1.5

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2	13.717	5.56%	1.288	4.2028
0	**-1.5**	**13.638**	**5.53%**	**1.293**	**4.1988**
0	-1	13.760	5.58%	1.315	4.2535
0	-0.5	14.017	5.68%	1.361	4.3678
0	0	14.549	5.90%	1.444	4.5833
0	0.5	15.379	6.23%	1.587	4.9399
0	1	16.443	6.67%	1.839	5.4984
0	1.5	17.737	7.25%	2.310	6.4013
0	2	20.678	8.65%	3.273	8.2267
0.5	-1.5	11.539	4.61%	1.044	3.4712
0.5	**-1**	**11.559**	**4.64%**	**1.041**	**3.4683**
0.5	-0.5	11.718	4.72%	1.042	3.4938
0.5	0	11.980	4.84%	1.049	3.5449
0.5	0.5	12.283	4.98%	1.065	3.6168
0.5	1	12.627	5.14%	1.094	3.7161
0.5	1.5	13.014	5.33%	1.141	3.8531
0.5	2	13.445	5.54%	1.217	4.0455
1	-1	10.642	4.12%	1.041	3.3290
1	**-0.5**	**10.692**	**4.17%**	**1.033**	**3.3231**
1	0	10.798	4.23%	1.025	3.3272
1	0.5	10.971	4.33%	1.019	3.3443
1	1	11.169	4.43%	1.017	3.3700
1	1.5	11.423	4.56%	1.019	3.4119
1	2	11.710	4.71%	1.028	3.4699
1.5	0	10.326	3.91%	1.083	3.3435
1.5	0.5	10.400	3.96%	1.069	3.3341
1.5	**1**	**10.478**	**4.01%**	**1.056**	**3.3257**
1.5	1.5	10.617	4.10%	1.043	3.3283

1.5	2	10.787	4.19%	1.034	3.3393
2	0	10.271	3.80%	1.174	3.4725
2	0.5	10.333	3.84%	1.156	3.4566
2	1	10.404	3.89%	1.138	3.4404
2	1.5	10.480	3.94%	1.119	3.4241
2	2	10.559	3.99%	1.100	3.4085
2	2.5	10.643	4.05%	1.083	3.3950
2	3	10.730	4.12%	1.068	3.3851
2	**3.5**	**10.821**	**4.18%**	**1.056**	**3.3810**
2	4	10.970	4.28%	1.050	3.3938

Table 12: The Model Performance of GMBM with K=2

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2	13.813	5.56%	1.313	4.2583
0	**-1.5**	**13.667**	**5.51%**	**1.316**	**4.2415**
0	-1	13.803	5.56%	1.337	4.2964
0	-0.5	14.015	5.65%	1.382	4.4014
0	0	14.564	5.86%	1.463	4.6157
0	0.5	15.373	6.18%	1.600	4.9588
0	1	16.366	6.58%	1.835	5.4794
0	1.5	17.553	7.11%	2.260	6.2987
0	2	20.067	8.30%	3.091	7.8758
0.5	-1.5	11.538	4.59%	1.043	3.4683
0.5	**-1**	**11.564**	**4.62%**	**1.040**	**3.4681**
0.5	-0.5	11.715	4.70%	1.043	3.4948
0.5	0	11.983	4.82%	1.052	3.5508
0.5	0.5	12.290	4.96%	1.072	3.6292
0.5	1	12.637	5.12%	1.105	3.7364
0.5	1.5	13.023	5.30%	1.157	3.8821
0.5	2	13.448	5.51%	1.240	4.0832
1	-1	10.614	4.10%	1.042	3.3248
1	**-0.5**	**10.666**	**4.14%**	**1.032**	**3.3178**
1	0	10.767	4.20%	1.024	3.3199
1	0.5	10.957	4.30%	1.017	3.3390
1	1	11.192	4.42%	1.015	3.3705
1	1.5	11.458	4.56%	1.018	3.4156
1	2	11.756	4.71%	1.029	3.4783
1.5	0	10.304	3.89%	1.087	3.3470
1.5	0.5	10.376	3.94%	1.072	3.3350
1.5	**1**	**10.454**	**3.99%**	**1.057**	**3.3241**
1.5	1.5	10.596	4.07%	1.043	3.3250

1.5	2	10.768	4.17%	1.032	3.3342
2	0	10.257	3.78%	1.184	3.4855
2	0.5	10.320	3.82%	1.166	3.4683
2	1	10.390	3.87%	1.146	3.4503
2	1.5	10.464	3.92%	1.125	3.4317
2	2	10.542	3.97%	1.105	3.4135
2	2.5	10.624	4.03%	1.086	3.3970
2	3	10.709	4.09%	1.069	3.3838
2	**3.5**	**10.797**	**4.16%**	**1.056**	**3.3762**
2	4	10.955	4.26%	1.047	3.3873

Table 13: The Model Performance of GMBM with K=2.5

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2	13.931	5.57%	1.342	4.3236
0	**-1.5**	**13.770**	**5.52%**	**1.343**	**4.3003**
0	-1	13.840	5.54%	1.362	4.3415
0	-0.5	14.040	5.62%	1.405	4.4413
0	0	14.575	5.82%	1.482	4.6474
0	0.5	15.333	6.12%	1.611	4.9699
0	1	16.259	6.49%	1.827	5.4508
0	1.5	17.354	6.97%	2.208	6.1896
0	2	19.455	7.95%	2.922	7.5395
0.5	-2	11.570	4.57%	1.050	3.4852
0.5	**-1.5**	**11.536**	**4.57%**	**1.045**	**3.4715**
0.5	-1	11.571	4.60%	1.044	3.4749
0.5	-0.5	11.712	4.68%	1.048	3.5037
0.5	0	11.986	4.80%	1.061	3.5655
0.5	0.5	12.298	4.94%	1.084	3.6510
0.5	1	12.647	5.10%	1.121	3.7658
0.5	1.5	13.030	5.28%	1.179	3.9188
0.5	2	13.477	5.49%	1.266	4.1300
1	-1	10.584	4.07%	1.044	3.3236
1	**-0.5**	**10.639**	**4.11%**	**1.034**	**3.3159**
1	0	10.761	4.18%	1.025	3.3207
1	0.5	10.974	4.29%	1.019	3.3435
1	1	11.218	4.41%	1.017	3.3779
1	1.5	11.495	4.55%	1.022	3.4273
1	2	11.804	4.70%	1.035	3.4961
1.5	0	10.288	3.87%	1.093	3.3531
1.5	0.5	10.353	3.91%	1.076	3.3381
1.5	1	10.430	3.97%	1.060	3.3252

1.5	**1.5**	**10.573**	**4.05%**	**1.045**	**3.3245**
1.5	2	10.748	4.15%	1.033	3.3326
2	0	10.252	3.76%	1.196	3.5009
2	0.5	10.308	3.80%	1.176	3.4811
2	1	10.377	3.85%	1.155	3.4613
2	1.5	10.449	3.90%	1.133	3.4406
2	2	10.526	3.96%	1.111	3.4200
2	2.5	10.605	4.01%	1.091	3.4008
2	3	10.688	4.07%	1.072	3.3847
2	**3.5**	**10.773**	**4.13%**	**1.057**	**3.3741**
2	4	10.940	4.24%	1.047	3.3841

Table 14: The Model Performance of GMBM with K=3

p	q	wab	$wapb$	$wChi$	$\sqrt{wab*wChi}$
0	-2	14.081	5.59%	1.373	4.3971
0	**-1.5**	**13.870**	**5.52%**	**1.371**	**4.3603**
0	-1	13.884	5.53%	1.387	4.3882
0	-0.5	14.110	5.61%	1.427	4.4879
0	0	14.553	5.78%	1.500	4.6729
0	0.5	15.266	6.05%	1.621	4.9744
0	1	16.133	6.40%	1.818	5.4164
0	1.5	17.149	6.84%	2.156	6.0804
0	2	18.860	7.63%	2.769	7.2268
0.5	-2	11.563	4.55%	1.054	3.4916
0.5	**-1.5**	**11.537**	**4.55%**	**1.050**	**3.4813**
0.5	-1	11.581	4.58%	1.051	3.4892
0.5	-0.5	11.711	4.65%	1.058	3.5208
0.5	0	11.992	4.78%	1.074	3.5892
0.5	0.5	12.308	4.92%	1.101	3.6818
0.5	1	12.659	5.08%	1.143	3.8036
0.5	1.5	13.038	5.25%	1.204	3.9619
0.5	2	13.498	5.46%	1.293	4.1783
1	-1	10.560	4.04%	1.048	3.3269
1	**-0.5**	**10.645**	**4.10%**	**1.038**	**3.3236**
1	0	10.771	4.17%	1.029	3.3291
1	0.5	10.992	4.28%	1.024	3.3546
1	1	11.246	4.41%	1.024	3.3930
1	1.5	11.533	4.55%	1.031	3.4480
1	2	11.851	4.70%	1.048	3.5238
1.5	0	10.272	3.85%	1.100	3.3613
1.5	0.5	10.330	3.89%	1.082	3.3435
1.5	1	10.407	3.94%	1.065	3.3291

1.5	**1.5**	**10.557**	**4.03%**	**1.049**	**3.3285**
1.5	2	10.765	4.14%	1.037	3.3413
2	0	10.247	3.75%	1.207	3.5170
2	0.5	10.296	3.79%	1.186	3.4947
2	1	10.363	3.83%	1.164	3.4733
2	1.5	10.435	3.88%	1.141	3.4508
2	2	10.509	3.94%	1.118	3.4282
2	2.5	10.587	3.99%	1.096	3.4066
2	3	10.667	4.05%	1.076	3.3882
2	**3.5**	**10.750**	**4.11%**	**1.060**	**3.3753**
2	4	10.930	4.22%	1.049	3.3859

Appendix 3: Numerical Iterative Process of GMBM

Table 15: Numerical Iterations for Multiplicative Gamma GMBM Factors Using Average Severity as Base

Iteration	Base	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	241.46	1.203551	1.207636	1.154385	1.123670	0.890524	0.970979	0.953411	0.921830	1.393434	1.073693	0.887450	0.854173
2	241.46	1.239832	1.234406	1.144728	1.097253	0.883394	0.955745	0.969955	0.948637	1.398901	1.075486	0.886537	0.850980
3	241.46	1.240574	1.234754	1.144648	1.096890	0.883231	0.955539	0.970167	0.949079	1.398980	1.075513	0.886525	0.850929
4	241.46	1.240585	1.234759	1.144647	1.096885	0.883229	0.955536	0.970170	0.949086	1.398981	1.075513	0.886525	0.850928

Table 16: Numerical Iterations for Multiplicative Gamma GMBM Factors Using 60+ and Pleasure as the Base

Iteration	Base	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	190.126	1.306	1.310	1.252	1.219	0.966	1.053	1.034	1.000	1.631	1.257	1.039	1.000
2	194.924	1.307	1.301	1.207	1.157	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000
3	195.003	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000
4	195.004	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000

Table 17: Numerical Iterations for Multiplicative Gamma GLM Coefficients Using 60+ and Pleasure as the Base

Iteration	Intercept	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	5.2710091	0.2447301	0.2563505	0.1870363	0.1454066	-0.0752765	0.0065328	0.0224012	0	0.4944179	0.2360242	0.0430371	0
2	5.2729277	0.2683345	0.2629985	0.1872796	0.1447046	-0.0720638	0.0068159	0.0219726	0	0.497298	0.2342915	0.0411756	0
3	5.2730182	0.2678326	0.263122	0.1873523	0.1447304	-0.0719202	0.0067721	0.0219717	0	0.4971672	0.234231	0.0409872	0
4	5.2730202	0.2678398	0.2631314	0.1873525	0.14473	-0.0719157	0.0067735	0.0219717	0	0.4971718	0.2342254	0.040982	0

Table 18: Numerical Iterations for Multiplicative Gamma GLM Factors Using 60+ and Pleasure as the Base

Iteration	Base	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	194.612	1.277	1.292	1.206	1.157	0.927	1.007	1.023	1.000	1.640	1.266	1.044	1.000
2	194.986	1.308	1.301	1.206	1.156	0.930	1.007	1.022	1.000	1.644	1.264	1.042	1.000
3	195.004	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000
4	195.004	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000

Table 19: Numerical Iterations for Additive Factors of Gamma GMBM Using Average Severity as the Base

Iteration	Base	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	241.46	1.203551	1.207636	1.154385	1.123670	0.890524	0.970979	0.953411	0.921830	0.393760	0.072094	-0.111752	-0.145133
2	241.46	1.247269	1.219244	1.138179	1.101277	0.875751	0.958654	0.972661	0.955564	0.398400	0.074096	-0.113076	-0.149284
3	241.46	1.248066	1.219512	1.137964	1.100911	0.875520	0.958409	0.972929	0.956185	0.398478	0.074130	-0.113097	-0.149360
4	241.46	1.248079	1.219517	1.137960	1.100904	0.875516	0.958405	0.972934	0.956196	0.398479	0.074131	-0.113097	-0.149361
5	241.46	1.248080	1.219517	1.137960	1.100904	0.875516	0.958405	0.972934	0.956196	0.398479	0.074131	-0.113097	-0.149361

Table 20: Numerical Iterations for Additive Dollar Values of Gamma GMBM Using Age 60+ and Pleasure as Base

Iteration	Base	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	187.5412	68.0244	69.0107	56.1527	48.7365	-7.5591	11.8675	7.6257	0.0000	130.1212	52.4515	8.0601	0.0000
2	194.6844	70.4351	63.6680	44.0942	35.1839	-19.2717	0.7462	4.1282	0.0000	132.2437	53.9373	8.7428	0.0000
3	194.8161	70.4774	63.5829	43.8922	34.9454	-19.4776	0.5369	4.0430	0.0000	132.2809	53.9639	8.7561	0.0000
4	194.8184	70.4781	63.5815	43.8887	34.9412	-19.4811	0.5333	4.0414	0.0000	132.2815	53.9644	8.7563	0.0000
5	194.8185	70.4781	63.5814	43.8887	34.9412	-19.4812	0.5332	4.0414	0.0000	132.2815	53.9644	8.7563	0.0000

Table 21: Additive Normal GLM Coefficients Using 60+ and Pleasure as the Base

GLM	Intercept	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
Coefficient	194.8185	70.4781	63.5814	43.8887	34.9412	-19.4812	0.5332	4.0414	0.0000	132.2815	53.9644	8.7563	0.0000

Biographies of Authors

Luyang Fu is a pricing actuary in Grange Mutual Insurance Company. Mr. Fu received his Ph.D. in agricultural and consumer economics and master in finance from University of Illinois at Urbana-Champaign.

Cheng-sheng Peter Wu, FCAS, ASA, MAAA, is a Director in the Advanced Quantitative Services practice of Deloitte & Touche's Actuarial and Insurance Consulting Group. He is based in the Los Angeles, CA office. Mr. Wu received his Masters degrees in chemical engineering and statistics from the Pennsylvania State University. Mr. Wu has published several papers in automotive engineering, tribology (lubrication engineering), statistics, and actuarial science.

Table 9: The Model Performance of GMBM with K=0.5

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2.5	13.747	5.63%	1.262	4.1651
0	**-2**	**13.545**	**5.56%**	**1.255**	**4.1225**
0	-1.5	13.588	5.58%	1.260	4.1384
0	-1	13.770	5.65%	1.283	4.2024
0	-0.5	14.050	5.77%	1.328	4.3198
0	0	14.618	6.01%	1.411	4.5418
0	0.5	15.337	6.32%	1.559	4.8894
0	1	16.467	6.80%	1.832	5.4920
0	1.5	18.211	7.58%	2.382	6.5860
0	2	21.788	9.33%	3.642	8.9078
0.5	-1.5	11.593	4.67%	1.056	3.4992
0.5	**-1**	**11.584**	**4.69%**	**1.052**	**3.4912**
0.5	-0.5	11.735	4.76%	1.052	3.5131
0.5	0	11.972	4.88%	1.056	3.5559
0.5	0.5	12.261	5.01%	1.068	3.6180
0.5	1	12.594	5.17%	1.089	3.7036
0.5	1.5	12.974	5.35%	1.126	3.8221
0.5	2	13.410	5.57%	1.187	3.9900
1	-1	10.696	4.18%	1.046	3.3453
1	**-0.5**	**10.739**	**4.22%**	**1.040**	**3.3420**
1	0	10.851	4.28%	1.035	3.3506
1	0.5	11.018	4.37%	1.031	3.3699
1	1	11.208	4.47%	1.029	3.3967
1	1.5	11.422	4.59%	1.032	3.4330
1	2	11.658	4.72%	1.040	3.4818
1.5	0	10.376	3.96%	1.077	3.3435
1.5	0.5	10.452	4.01%	1.067	3.3391
1.5	**1**	**10.531**	**4.06%**	**1.057**	**3.3359**
1.5	1.5	10.657	4.14%	1.048	3.3419
1.5	2	10.820	4.23%	1.042	3.3572
2	0	10.310	3.84%	1.155	3.4515
2	0.5	10.374	3.88%	1.140	3.4389
2	1	10.443	3.93%	1.124	3.4262
2	1.5	10.518	3.98%	1.108	3.4143
2	2	10.600	4.04%	1.093	3.4040
2	2.5	10.688	4.10%	1.080	3.3969
2	**3**	**10.781**	**4.16%**	**1.069**	**3.3943**
2	3.5	10.877	4.23%	1.062	3.3982

Table 10: The Model Performance of GMBM with K=1

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2.5	13.873	5.65%	1.276	4.2068
0	**-2**	**13.631**	**5.56%**	**1.268**	**4.1575**
0	-1.5	13.614	5.56%	1.274	4.1642
0	-1	13.753	5.61%	1.296	4.2221
0	-0.5	14.018	5.72%	1.343	4.3381
0	0	14.588	5.96%	1.426	4.5612
0	0.5	15.344	6.27%	1.573	4.9128
0	1	16.480	6.74%	1.838	5.5038
0	1.5	18.011	7.43%	2.353	6.5095
0	2	21.263	9.00%	3.460	8.5778
0.5	-1.5	11.563	4.64%	1.049	3.4825
0.5	**-1**	**11.554**	**4.66%**	**1.045**	**3.4746**
0.5	-0.5	11.722	4.74%	1.045	3.4998
0.5	0	11.977	4.86%	1.050	3.5470
0.5	0.5	12.274	5.00%	1.064	3.6134
0.5	1	12.613	5.16%	1.089	3.7054
0.5	1.5	12.998	5.34%	1.130	3.8331
0.5	2	13.434	5.56%	1.199	4.0139
1	-1	10.669	4.15%	1.043	3.3358
1	**-0.5**	**10.716**	**4.19%**	**1.036**	**3.3313**
1	0	10.826	4.26%	1.029	3.3376
1	0.5	10.996	4.35%	1.024	3.3556
1	1	11.190	4.45%	1.022	3.3815
1	1.5	11.408	4.57%	1.024	3.4178
1	2	11.664	4.70%	1.032	3.4696
1.5	0	10.350	3.94%	1.079	3.3423
1.5	0.5	10.424	3.99%	1.067	3.3353
1.5	**1**	**10.503**	**4.04%**	**1.055**	**3.3295**
1.5	1.5	10.637	4.12%	1.045	3.3340
1.5	2	10.804	4.21%	1.037	3.3471
2	0	10.290	3.82%	1.164	3.4614
2	0.5	10.353	3.86%	1.148	3.4471
2	1	10.422	3.91%	1.130	3.4323
2	1.5	10.495	3.96%	1.113	3.4179
2	2	10.577	4.01%	1.096	3.4051
2	2.5	10.663	4.07%	1.081	3.3947
2	3	10.753	4.14%	1.068	3.3883
2	**3.5**	**10.846**	**4.21%**	**1.058**	**3.3881**
2	4	10.983	4.30%	1.054	3.4028

Table 11: The Model Performance of GMBM with K=1.5

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2	13.717	5.56%	1.288	4.2028
0	**-1.5**	**13.638**	**5.53%**	**1.293**	**4.1988**
0	-1	13.760	5.58%	1.315	4.2535
0	-0.5	14.017	5.68%	1.361	4.3678
0	0	14.549	5.90%	1.444	4.5833
0	0.5	15.379	6.23%	1.587	4.9399
0	1	16.443	6.67%	1.839	5.4984
0	1.5	17.737	7.25%	2.310	6.4013
0	2	20.678	8.65%	3.273	8.2267
0.5	-1.5	11.539	4.61%	1.044	3.4712
0.5	**-1**	**11.559**	**4.64%**	**1.041**	**3.4683**
0.5	-0.5	11.718	4.72%	1.042	3.4938
0.5	0	11.980	4.84%	1.049	3.5449
0.5	0.5	12.283	4.98%	1.065	3.6168
0.5	1	12.627	5.14%	1.094	3.7161
0.5	1.5	13.014	5.33%	1.141	3.8531
0.5	2	13.445	5.54%	1.217	4.0455
1	-1	10.642	4.12%	1.041	3.3290
1	**-0.5**	**10.692**	**4.17%**	**1.033**	**3.3231**
1	0	10.798	4.23%	1.025	3.3272
1	0.5	10.971	4.33%	1.019	3.3443
1	1	11.169	4.43%	1.017	3.3700
1	1.5	11.423	4.56%	1.019	3.4119
1	2	11.710	4.71%	1.028	3.4699
1.5	0	10.326	3.91%	1.083	3.3435
1.5	0.5	10.400	3.96%	1.069	3.3341
1.5	**1**	**10.478**	**4.01%**	**1.056**	**3.3257**
1.5	1.5	10.617	4.10%	1.043	3.3283
1.5	2	10.787	4.19%	1.034	3.3393
2	0	10.271	3.80%	1.174	3.4725
2	0.5	10.333	3.84%	1.156	3.4566
2	1	10.404	3.89%	1.138	3.4404
2	1.5	10.480	3.94%	1.119	3.4241
2	2	10.559	3.99%	1.100	3.4085
2	2.5	10.643	4.05%	1.083	3.3950
2	3	10.730	4.12%	1.068	3.3851
2	**3.5**	**10.821**	**4.18%**	**1.056**	**3.3810**
2	4	10.970	4.28%	1.050	3.3938

Table 12: The Model Performance of GMBM with K=2

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2	13.813	5.56%	1.313	4.2583
0	**-1.5**	**13.667**	**5.51%**	**1.316**	**4.2415**
0	-1	13.803	5.56%	1.337	4.2964
0	-0.5	14.015	5.65%	1.382	4.4014
0	0	14.564	5.86%	1.463	4.6157
0	0.5	15.373	6.18%	1.600	4.9588
0	1	16.366	6.58%	1.835	5.4794
0	1.5	17.553	7.11%	2.260	6.2987
0	2	20.067	8.30%	3.091	7.8758
0.5	-1.5	11.538	4.59%	1.043	3.4683
0.5	**-1**	**11.564**	**4.62%**	**1.040**	**3.4681**
0.5	-0.5	11.715	4.70%	1.043	3.4948
0.5	0	11.983	4.82%	1.052	3.5508
0.5	0.5	12.290	4.96%	1.072	3.6292
0.5	1	12.637	5.12%	1.105	3.7364
0.5	1.5	13.023	5.30%	1.157	3.8821
0.5	2	13.448	5.51%	1.240	4.0832
1	-1	10.614	4.10%	1.042	3.3248
1	**-0.5**	**10.666**	**4.14%**	**1.032**	**3.3178**
1	0	10.767	4.20%	1.024	3.3199
1	0.5	10.957	4.30%	1.017	3.3390
1	1	11.192	4.42%	1.015	3.3705
1	1.5	11.458	4.56%	1.018	3.4156
1	2	11.756	4.71%	1.029	3.4783
1.5	0	10.304	3.89%	1.087	3.3470
1.5	0.5	10.376	3.94%	1.072	3.3350
1.5	**1**	**10.454**	**3.99%**	**1.057**	**3.3241**
1.5	1.5	10.596	4.07%	1.043	3.3250
1.5	2	10.768	4.17%	1.032	3.3342
2	0	10.257	3.78%	1.184	3.4855
2	0.5	10.320	3.82%	1.166	3.4683
2	1	10.390	3.87%	1.146	3.4503
2	1.5	10.464	3.92%	1.125	3.4317
2	2	10.542	3.97%	1.105	3.4135
2	2.5	10.624	4.03%	1.086	3.3970
2	3	10.709	4.09%	1.069	3.3838
2	**3.5**	**10.797**	**4.16%**	**1.056**	**3.3762**
2	4	10.955	4.26%	1.047	3.3873

Table 13: The Model Performance of GMBM with K=2.5

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2	13.931	5.57%	1.342	4.3236
0	**-1.5**	**13.770**	**5.52%**	**1.343**	**4.3003**
0	-1	13.840	5.54%	1.362	4.3415
0	-0.5	14.040	5.62%	1.405	4.4413
0	0	14.575	5.82%	1.482	4.6474
0	0.5	15.333	6.12%	1.611	4.9699
0	1	16.259	6.49%	1.827	5.4508
0	1.5	17.354	6.97%	2.208	6.1896
0	2	19.455	7.95%	2.922	7.5395
0.5	-2	11.570	4.57%	1.050	3.4852
0.5	**-1.5**	**11.536**	**4.57%**	**1.045**	**3.4715**
0.5	-1	11.571	4.60%	1.044	3.4749
0.5	-0.5	11.712	4.68%	1.048	3.5037
0.5	0	11.986	4.80%	1.061	3.5655
0.5	0.5	12.298	4.94%	1.084	3.6510
0.5	1	12.647	5.10%	1.121	3.7658
0.5	1.5	13.030	5.28%	1.179	3.9188
0.5	2	13.477	5.49%	1.266	4.1300
1	-1	10.584	4.07%	1.044	3.3236
<u>**1**</u>	<u>**-0.5**</u>	<u>**10.639**</u>	<u>**4.11%**</u>	<u>**1.034**</u>	<u>**3.3159**</u>
1	0	10.761	4.18%	1.025	3.3207
1	0.5	10.974	4.29%	1.019	3.3435
1	1	11.218	4.41%	1.017	3.3779
1	1.5	11.495	4.55%	1.022	3.4273
1	2	11.804	4.70%	1.035	3.4961
1.5	0	10.288	3.87%	1.093	3.3531
1.5	0.5	10.353	3.91%	1.076	3.3381
1.5	1	10.430	3.97%	1.060	3.3252
1.5	**1.5**	**10.573**	**4.05%**	**1.045**	**3.3245**
1.5	2	10.748	4.15%	1.033	3.3326
2	0	10.252	3.76%	1.196	3.5009
2	0.5	10.308	3.80%	1.176	3.4811
2	1	10.377	3.85%	1.155	3.4613
2	1.5	10.449	3.90%	1.133	3.4406
2	2	10.526	3.96%	1.111	3.4200
2	2.5	10.605	4.01%	1.091	3.4008
2	3	10.688	4.07%	1.072	3.3847
2	**3.5**	**10.773**	**4.13%**	**1.057**	**3.3741**
2	4	10.940	4.24%	1.047	3.3841

Table 14: The Model Performance of GMBM with K=3

p	q	wab	$wapb$	$wChi$	$\sqrt{wab * wChi}$
0	-2	14.081	5.59%	1.373	4.3971
0	**-1.5**	**13.870**	**5.52%**	**1.371**	**4.3603**
0	-1	13.884	5.53%	1.387	4.3882
0	-0.5	14.110	5.61%	1.427	4.4879
0	0	14.553	5.78%	1.500	4.6729
0	0.5	15.266	6.05%	1.621	4.9744
0	1	16.133	6.40%	1.818	5.4164
0	1.5	17.149	6.84%	2.156	6.0804
0	2	18.860	7.63%	2.769	7.2268
0.5	-2	11.563	4.55%	1.054	3.4916
0.5	**-1.5**	**11.537**	**4.55%**	**1.050**	**3.4813**
0.5	-1	11.581	4.58%	1.051	3.4892
0.5	-0.5	11.711	4.65%	1.058	3.5208
0.5	0	11.992	4.78%	1.074	3.5892
0.5	0.5	12.308	4.92%	1.101	3.6818
0.5	1	12.659	5.08%	1.143	3.8036
0.5	1.5	13.038	5.25%	1.204	3.9619
0.5	2	13.498	5.46%	1.293	4.1783
1	-1	10.560	4.04%	1.048	3.3269
1	**-0.5**	**10.645**	**4.10%**	**1.038**	**3.3236**
1	0	10.771	4.17%	1.029	3.3291
1	0.5	10.992	4.28%	1.024	3.3546
1	1	11.246	4.41%	1.024	3.3930
1	1.5	11.533	4.55%	1.031	3.4480
1	2	11.851	4.70%	1.048	3.5238
1.5	0	10.272	3.85%	1.100	3.3613
1.5	0.5	10.330	3.89%	1.082	3.3435
1.5	1	10.407	3.94%	1.065	3.3291
1.5	**1.5**	**10.557**	**4.03%**	**1.049**	**3.3285**
1.5	2	10.765	4.14%	1.037	3.3413
2	0	10.247	3.75%	1.207	3.5170
2	0.5	10.296	3.79%	1.186	3.4947
2	1	10.363	3.83%	1.164	3.4733
2	1.5	10.435	3.88%	1.141	3.4508
2	2	10.509	3.94%	1.118	3.4282
2	2.5	10.587	3.99%	1.096	3.4066
2	3	10.667	4.05%	1.076	3.3882
2	**3.5**	**10.750**	**4.11%**	**1.060**	**3.3753**
2	4	10.930	4.22%	1.049	3.3859

Appendix 3: Numerical Iterative Process of GMBM

Table 15: Numerical Iterations for Multiplicative Gamma GMBM Factors Using Average Severity as Base

Iteration	Base	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	241.46	1.203551	1.207636	1.154385	1.123670	0.890524	0.970979	0.953411	0.921830	1.393434	1.073693	0.887450	0.854173
2	241.46	1.239832	1.234406	1.144728	1.097253	0.883394	0.955745	0.969955	0.948637	1.398901	1.075486	0.886537	0.850980
3	241.46	1.240574	1.234754	1.144648	1.096890	0.883231	0.955539	0.970167	0.949079	1.398980	1.075513	0.886525	0.850929
4	241.46	1.240585	1.234759	1.144647	1.096885	0.883229	0.955536	0.970170	0.949086	1.398981	1.075513	0.886525	0.850928

Table 16: Numerical Iterations for Multiplicative Gamma GMBM Factors Using 60+ and Pleasure as the Base

Iteration	Base	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	190.126	1.306	1.310	1.252	1.219	0.966	1.053	1.034	1.000	1.631	1.257	1.039	1.000
2	194.924	1.307	1.301	1.207	1.157	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000
3	195.003	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000
4	195.004	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000

Table 17: Numerical Iterations for Multiplicative Gamma GLM Coefficients Using 60+ and Pleasure as the Base

Iteration	Intercept	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	5.2710091	0.2447301	0.2563505	0.1870363	0.1454066	-0.0752765	0.0065328	0.0224012	0	0.4944179	0.2360242	0.0430371	0
2	5.2729277	0.2683345	0.2629985	0.1872796	0.1447046	-0.0720638	0.0068159	0.0219726	0	0.497298	0.2342915	0.0411756	0
3	5.2730182	0.2678326	0.263122	0.1873523	0.1447304	-0.0719202	0.0067721	0.0219717	0	0.4971672	0.234231	0.0409872	0
4	5.2730202	0.2678398	0.2631314	0.1873525	0.14473	-0.0719157	0.0067735	0.0219717	0	0.4971718	0.2342254	0.040982	0

Table 18: Numerical Iterations for Multiplicative Gamma GLM Factors Using 60+ and Pleasure as the Base

Iteration	Base	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	194.612	1.277	1.292	1.206	1.157	0.927	1.007	1.023	1.000	1.640	1.266	1.044	1.000
2	194.986	1.308	1.301	1.206	1.156	0.930	1.007	1.022	1.000	1.644	1.264	1.042	1.000
3	195.004	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000
4	195.004	1.307	1.301	1.206	1.156	0.931	1.007	1.022	1.000	1.644	1.264	1.042	1.000

Table 19: Numerical Iterations for Additive Factors of Gamma GMBM Using Average Severity as the Base

Iteration	Base	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	241.46	1.203551	1.207636	1.154385	1.123670	0.890524	0.970979	0.953411	0.921830	0.393760	0.072094	-0.111752	-0.145133
2	241.46	1.247269	1.219244	1.138179	1.101277	0.875751	0.958654	0.972661	0.955564	0.398400	0.074096	-0.113076	-0.149284
3	241.46	1.248066	1.219512	1.137964	1.100911	0.875520	0.958409	0.972929	0.956185	0.398478	0.074130	-0.113097	-0.149360
4	241.46	1.248079	1.219517	1.137960	1.100904	0.875516	0.958405	0.972934	0.956196	0.398479	0.074131	-0.113097	-0.149361
5	241.46	1.248080	1.219517	1.137960	1.100904	0.875516	0.958405	0.972934	0.956196	0.398479	0.074131	-0.113097	-0.149361

Table 20: Numerical Iterations for Additive Dollar Values of Gamma GMBM Using Age 60+ and Pleasure as Base

Iteration	Base	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
1	187.5412	68.0244	69.0107	56.1527	48.7365	-7.5591	11.8675	7.6257	0.0000	130.1212	52.4515	8.0601	0.0000
2	194.6844	70.4351	63.6680	44.0942	35.1839	-19.2717	0.7462	4.1282	0.0000	132.2437	53.9373	8.7428	0.0000
3	194.8161	70.4774	63.5829	43.8922	34.9454	-19.4776	0.5369	4.0430	0.0000	132.2809	53.9639	8.7561	0.0000
4	194.8184	70.4781	63.5815	43.8887	34.9412	-19.4811	0.5333	4.0414	0.0000	132.2815	53.9644	8.7563	0.0000
5	194.8185	70.4781	63.5814	43.8887	34.9412	-19.4812	0.5332	4.0414	0.0000	132.2815	53.9644	8.7563	0.0000

Table 21: Additive Normal GLM Coefficients Using 60+ and Pleasure as the Base

GLM	Intercept	17-20	21-24	25-29	30-34	35-39	40-49	50-59	60+	Business	DTW Long	DTW Short	Pleasure
Coefficient	194.8185	70.4781	63.5814	43.8887	34.9412	-19.4812	0.5332	4.0414	0.0000	132.2815	53.9644	8.7563	0.0000

Biographies of Authors

Luyang Fu is a pricing actuary in Grange Mutual Insurance Company. Mr. Fu received his Ph.D. in agricultural and consumer economics and master in finance from University of Illinois at Urbana-Champaign.

Cheng-sheng Peter Wu, FCAS, ASA, MAAA, is a Director in the Advanced Quantitative Services practice of Deloitte & Touche's Actuarial and Insurance Consulting Group. He is based in the Los Angeles, CA office. Mr. Wu received his Masters degrees in chemical engineering and statistics from the Pennsylvania State University. Mr. Wu has published several papers in automotive engineering, tribology (lubrication engineering), statistics, and actuarial science.

Discussion of Generalized Minimum Bias Models

by Luyang Fu, Ph. D. and Cheng-sheng Peter Wu, FCAS, ASA, MAAA

Discussion by Stephen J. Mildenhall, FCAS, ASA, MAAA

Abtract: Fu and Wu have presented three generalizations of the minimum bias model iterations and demonstrated the impact these generalizations have on fitted parameters. This discussion explains how their generalized minimum bias models correspond to generaliezd linear models.

Fu and Wu's paper introduces 2- and 3-parameter Generalized Minimum Bias Models (GMBMs) which the authors claim extend Generalized Linear Models (GLMs). The GMBM depends on three parameters p, q and k, and is specified by the iterative scheme

$$x_i^k = \frac{\sum_j w_{ij}^p r_{ij}^k y_j^{q-k}}{\sum_j w_{ij}^p y_j^q} \tag{1}$$

where the w are prior weights, r the observations and x, y the parameters. The model is multiplicative: the fitted expected value of r_{ij} is given by $x_i y_j$. We will call this model a GMBM(p, q, k). The authors provide numerous examples of fits with different p, q, k. Unfortunately, by examining the effect of the three parameters p, q and k we can show that every GMBM corresponds to a GLM, so the new models do not extend the existing statistical models. Statistical models and approaches should always be preferred to non-statistical minimum bias models.

The parameter p is used to adjust the weights used from w_{ij} to w_{ij}^p. This adjustment can also be made in a GLM; the weights can be chosen however the modeler likes so long as they are specified ahead of time.

The parameter k replaces the responses r_{ij} with r_{ij}^k. A model is then fitted to the new responses to get parameters x_i and y_j. Finally, these parameters are converted back to the scale of the original responses by taking kth roots. Again, this procedure carries over to GLMs. Prior to modeling, replace each r_{ij} with r_{ij}^k, fit the model, and then replace the resulting fit parameters x_i, y_j with $x_i^{1/k}$ and $y_j^{1/k}$ respectively.

The parameter q is the most interesting. Comparing Equation 12 in the paper with Mildenhall [1999, Equation 7.13] shows that a value of q corresponds to using a variance function $V(\mu) = \mu^{2-q}$ in the GLM. As discussed in Mildenhall [1999, Section 8] there is a whole family of exponential distributions with variance $V(\mu) = \mu^\zeta$ where μ is the mean. The correspondence between ζ and distributions is shown in the table below. The common special cases are the normal $\zeta = 0$,

Table 1: Variance Functions

ζ	Distribution
$\zeta < 0$	Extreme Stable
0	Normal
$0 < \zeta < 1$	Not Exponential Family
1	Poisson
$1 < \zeta < 2$	Tweedie
2	Gamma
$2 < \zeta < \infty, \zeta \neq 3$	Positive Stable
3	Inverse Gaussian

Poisson $\zeta = 1$, gamma $\zeta = 2$ and inverse Gaussian $\zeta = 3$. These families are discussed more in McCullagh and Nelder [1989], Jörgensen [1997] and Jörgensen [1987]

The table shows that when $\zeta \neq 0, 1, 2, 3$ and $\zeta \notin (0, 1)$ there is still an exponential family corresponding to the variance function $V(\mu) = \mu^\zeta$. However, these distributions do not have a closed form expression for their densities. It is still possible to fit a GLM using these densities because the basic form of the likelihood function is known from the fact the distributions are in the exponential family. As explained in Mildenhall [1999, Section 8] and McCullagh and Nelder [1989] the deviance of an individual observation r_i is

$$2w_i \int_\mu^{r_i} \frac{r_i - t}{V(t)} dt + \log(V(r_i)). \tag{2}$$

This quantity is called the extended quasi-likelihood. It can be computed given the the function V only. It does not need the whole density. When $\zeta \in (0, 1)$ Equation 2 still makes sense (r_i and μ are positive in the examples) and it can be used in the GLM algorithm. However, such a variance function does not correspond to an exponential family distribution. Thus it is possible to work with GLMs with arbitrary $\zeta = 2 - q$.

Putting all three of these adjustments together gives the following dictionary between GMBMs and GLMs. The parameters produced by a GMBM(p, q, k) correspond to the kth roots of the parameters produced by a GLM with log link and weights w^p applied to data r^k and variance function $V(\mu) = \mu^\zeta$ where $\zeta = 2 - q/k$. The relativities in the appendix of Fu and Wu can be produced by GLMs in this way. When $\zeta \in (0, 1)$, for example $k = 1, q = 1.5$, there is no exponential family distribution member, but the GLM iteratively re-weighted least squares method still converges to the same values as given in the paper.

The paper also claims that using the most recent evaluation of each parameter in the iterative process greatly speeds up convergence. Subsequent to completing Mildenhall [1999] I read in Golub and Loan [1996, Section 10.1.1] that for the basic linear additive model, the minimum bias iterations were discovered by Jacobi, and are sometimes called the Jacobi iterations. Golub and Loan [1996] also contains the same idea for improving convergence that the authors suggest. For the linear additive model it is called the Gauss-Seidel iteration. In terms of overall speed of computation, the re-weighted least squares approach is like a higher-dimensional version of the Newton-Raphson method. The Newton-Raphson method converges extremely quickly. As explained in Mildenhall [1999], the basic minimum bias method converges as powers of the largest eigenvalue of a certain matrix. It can converge much more slowly than the GLM method. The improved scheme is clearly faster than the original but it may not be as quick as the re-weighted least squares algorithm.

References

Stephen J. Mildenhall. A systematic relationship between minimum bias and generalized linear models. *Proceedings of the Casualty Actuarial Society*, LXXXVI:393–487, 1999.

P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, second edition, 1989.

Bent Jörgensen. *The Theory of Dispersion Models*. Chapman and Hall, 1997.

Bent Jörgensen. Exponential dispersion models. *J. R. Statist. Soc. B*, 49(2):127–162, 1987.

Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, third edition, 1996.

The Effect of Changing Exposure Levels on Calendar Year Loss Trends

Chris Styrsky, FCAS, MAAA

Abstract

This purpose of this paper is to illustrate the impact that changing exposure levels have on calendar year loss trends by creating a situation where the calendar year loss trends are inaccurate. The results show that the calendar year loss trends can be distorted significantly by exposure level changes, with the potential to affect rate levels if not accounted for.

However, the effect of changing exposure could be accounted for. The proposed method of data organization will allow the impact of changing exposure levels to be negated, allowing actuaries to set more precise rates.

Due to the significant impact that changing exposure levels can have on the calendar year loss trends, it would be beneficial to organize the data in a similar fashion to what is proposed in this paper. This would reduce the chance of increasing market share at an inadequate rate or decreasing market share with an excessive rate.

Keywords. Calendar year, loss trends, data organization, exposure level change.

1 INTRODUCTION

When pricing a product in a competitive market, a delicate balance is struck between acquiring market share and the rate of return. Missteps in setting a proper price can lead to either an overpriced product that few will purchase or an inadequate rate that many will purchase but will not result in a sufficient profit.

The insurance industry's difficulty is compounded by the fact that insurance companies do not know what the actual cost of the product is until some time after it is sold. This makes the ability to accurately forecast the price of insurance contracts of the utmost importance.

According to the Statement of Principles for Ratemaking, an actuary should consider data organization and trends when determining a rate. The choice of data organization often used in trend analysis is calendar year. This is done because of the responsiveness of calendar year data and that calendar year data is readily available. The calendar year loss trends are used as guidance for the actuary to project historical data to reflect loss cost differences over time. Without carefully considering what is driving the underlying data, however, a trend may be selected that will have one of two effects. Either the product will be so under-priced that the company's bottom line will be hurt, or that product will be so over-priced that it will be uncompetitive.

Unfortunately, calendar year data does have limitations. One of the underlying assumptions when using calendar year loss trends is that the book of business is relatively stable in size. This is often not a reasonable assumption, and, as a result, the calendar year loss trend will be a distorted reflection of reality.

The goal of this paper is to show how and why calendar year loss trends are distorted by changes in exposure levels and to propose an alternative method that eliminates the need to assume constant exposure level.

2 ANALYSIS USING CALENDAR YEAR LOSS TRENDS

One disadvantage of using calendar year data is the influence from multiple accident years within a single calendar year. This is particularly evident when calendar year data is used to calculate loss trends.

The following formulas are typically used to calculate calendar year paid frequency, severity, and pure premium for trending:

CY_X Paid Frequency = $(C_{0,12,X} + C_{12,24,X} + \ldots) / E_X$

CY_X Paid Severity = $(L_{0,12,X} + L_{12,24,X} + \ldots) / (C_{0,12,X} + C_{12,24,X} + \ldots)$

CY_X Paid Pure Premium = $(L_{0,12,X} + L_{12,24,X} + \ldots)/ E_X$

Where:

- CY_X = Calendar year X
- $C_{T,T+12,X}$ = # of claims paid during CY_X that were paid between T and T + 12 months after the claim occurred
- $L_{T,T+12,X}$ = $'s of paid losses during CY_X that were paid between T and T + 12 months after the claim occurred
- $S_{T,T+12,X}$ = The average paid severity of claims paid during CY_X between T and T + 12 months after the claim occurred
- E_X = Earned Exposures from calendar year X.

Graphically, the diagonal in Table 1 below represents the accident year X+1 paid claims. Accident year X+1 potentially contributes claims to calendar years X+1, X+2, X+3, and X+4.

TABLE 1

	Claim Payment Period			
Calendar Year	0-12	12-24	24-36	36-48
X	$C_{0,12,X}$	$C_{12,24,X}$	$C_{24,36,X}$	$C_{36,48,X}$
X+1	$C_{0,12,X+1}$	$C_{12,24,X+1}$	$C_{24,36,X+1}$	$C_{36,48,X+1}$
X+2	$C_{0,12,X+2}$	$C_{12,24,X+2}$	$C_{24,36,X+2}$	$C_{36,48,X+2}$
X+3	$C_{0,12,X+3}$	$C_{12,24,X+3}$	$C_{24,36,X+3}$	$C_{36,48,X+3}$
X+4	$C_{0,12,X+4}$	$C_{12,24,X+4}$	$C_{24,36,X+4}$	$C_{36,48,X+4}$

Let's take a look at a few examples to see how calendar year data is dependent on exposure level. We will use the following assumptions for each example:

- All policies are written on January 1st and are 12 month policies
- The ultimate claim frequency for every risk in existence is 0.20
- 50% of the ultimate claims are paid within 12 months of the date the policy was written, 30% between 12 and 24 months, and 20% between 24 and 36 months (no claims paid past 36 months)
- The claim payment pattern does not change over time
- During calendar year X+2, claims that were paid within 12 months of the date the policy was written were settled for $100, $200 for claims between 12 to 24 months, and $400 for claims between 24 to 36 months
- Annual inflation is 5% for all claims

2.1 No Exposure Level Change

The following chart contains the exposure for this example.

TABLE 2	
Calendar Year	Earned Exposures
X	100,000
X+1	100,000
X+2	100,000
X+3	100,000
X+4	100,000
X+5	100,000
X+6	100,000

Based on the exposure level:

TABLE 3					
	Calendar Year				
	Paid Claims				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	4,000[1]				
X+1	6,000	4,000			
X+2	10,000	6,000	4,000		
X+3		10,000	6,000	4,000	
X+4			10,000	6,000	4,000
X+5				10,000	6,000
X+6					10,000
All AY (total CY)	20,000[2]	20,000	20,000	20,000	20,000
CY Pd Freq	0.2	0.2	0.2	0.2	0.2
Change vs. Prior Year		0.0%	0.0%	0.0%	0.0%

[1] 4,000 = 0.04 * 100,000

[2] 0.20 = 20,000 / 100,000

In this example, the calendar year paid frequency is 0.2 for each year, resulting in a 0% trend.

The calendar year paid severity is:

TABLE 4					
	Calendar Year Paid Losses				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	$1,600,000[1]				
X+1	$1,200,000	$1,680,000[2]			
X+2	$1,000,000	$1,260,000	$1,764,000		
X+3		$1,050,000	$1,323,000	$1,852,200	
X+4			$1,102,500	$1,389,150	$1,944,810
X+5				$1,157,625	$1,458,608
X+6					$1,215,506
All AY (total CY)	$3,800,000	$3,990,000	$4,189,500	$4,398,975	$4,618,924
CY Pd Severity	$190.00[3]	$199.50	$209.48	$219.95	$230.95
Change vs. Prior Year		5.0%	5.0%	5.0%	5.0%

[1] $1,600,000 = 4000 * 400

[2] $1,680,000 = 4000 * 400 * 1.05

[3] $190.00 = 3,800,000 / 20,000

The resulting calendar year paid severity trend is 5%, which matches the inflation rate.

The calendar year paid pure premium is:

TABLE 5					
	Calendar Year Paid Losses				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	$1,600,000[1]				
X+1	$1,200,000	$1,680,000[2]			
X+2	$1,000,000	$1,260,000	$1,764,000		
X+3		$1,050,000	$1,323,000	$1,852,200	
X+4			$1,102,500	$1,389,150	$1,944,810
X+5				$1,157,625	$1,458,608
X+6					$1,215,506
All AY (total CY)	$3,800,000	$3,990,000	$4,189,500	$4,398,975	$4,618,924
CY Pd Pure Premium	$38.00[3]	$39.90	$41.90	$43.99	$46.19
Change vs. Prior Year		5.0%	5.0%	5.0%	5.0%

[1] $1,600,000 = 4000 * 400

[2] $1,680,000 = 4000 * 400 * 1.05

[3] $38.00 = 3,800,000 / 100,000

In this example, the calendar year pure premium trend is 5%, which equals (1 + pd freq trend) * (1 + pd sev trend) – 1.

2.2 Increasing Exposure Level

The following chart contains the exposure for this example.

TABLE 6	
Calendar Year	Earned Exposures
X	100,000
X+1	100,000
X+2	100,000
X+3	104,200
X+4	111,275
X+5	122,700
X+6	139,500

Based on this exposure level:

TABLE 7					
	Calendar Year				
	Paid Claims				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	4,000				
X+1	6,000	4,000			
X+2	10,000	6,000	4,000		
X+3		10,420	6,252	4,168	
X+4			11,128	6,677	4,451
X+5				12,270	7,362
X+6					13,950
All AY (total CY)	20,000	20,420	21,380	23,115	25,763
CY Pd Freq	0.2	0.1960	0.1921	0.1884	0.1847
Change vs. Prior Year		-2.0%	-2.0%	-2.0%	-2.0%

The result is important to note. One of the assumptions is that every exposure has an ultimate frequency of 0.2 (i.e. the paid frequency trend should be 0%), but based on using calendar year data a –2.0% paid frequency trend is measured.

The artificial trend is a mismatch between the numerator and the denominator of the formula used to calculate the calendar year paid frequency. The formula for CY_X Paid Frequency is $(C_{0,12,X} + C_{12,24,X} + \ldots) / E_x$. The numerator contains multiple accident years produced from different exposure levels (years X-1, X-2, …) while the denominator is the most recent calendar year exposures (year X). Since the numerator contains some claims that were produced by a different set of exposures, the possibility of a mismatch is possible unless the exposure levels in years X-1, X-2, … just happened to stay constant.

Therefore, the following observations can be made:

- The "true" paid frequency trend will not be captured with calendar year data unless the change in exposure level is the same from year to year
- If there is a constant non-zero change in exposure level the absolute paid frequency will not be accurate even though the trend is.

The calendar year paid severity is:

TABLE 8					
	Calendar Year				
	Paid Losses				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	$1,600,000				
X+1	$1,200,000	$1,680,000			
X+2	$1,000,000	$1,260,000	$1,764,000		
X+3		$1,094,100	$1,378,566	$1,929,992	
X+4			$1,226,807	$1,545,777	$2,164,087
X+5				$1,420,406	$1,789,711
X+6					$1,695,637
All AY (total CY)	$3,800,000	$4,034,100	$4,369,373	$4,896,175	$5,649,436
CY Pd Severity	$190.00	$197.56	$204.37	$211.82	$219.28
Change vs. Prior Year		4.0%	3.5%	3.6%	3.5%

The resulting calendar year paid severity trend is approximately 3.5%, well below the inflation rate of 5%.

The reason that the calendar year paid severity trend is different than the inflation rate is not that intuitive. The change in exposure level changes the distribution of each calendar year's claims by accident year. With increasing exposure level, the latest calendar year contains a higher percentage of paid claims from recent accident years (and those claims typically have a smaller severity). For example, in calendar year X+2 50% of the claims were from claims settled within 12 months of policy inception, 30% from 12-24 months, and 20% from 24-36. In calendar year X+3 the distribution was 51%, 29.4%, and 19.6%.

The calendar year paid pure premium is:

TABLE 9					
	Calendar Year Paid Losses				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	$1,600,000				
X+1	$1,200,000	$1,680,000			
X+2	$1,000,000	$1,260,000	$1,764,000		
X+3		$1,094,100	$1,378,566	$1,929,992	
X+4			$1,226,807	$1,545,777	$2,164,087
X+5				$1,420,406	$1,789,711
X+6					$1,695,637
All AY (total CY)	$3,800,000	$4,034,100	$4,369,373	$4,896,175	$5,649,436
CY Pd Pure Premium	$38.00	$38.71	$39.27	$39.90	$40.50
Change vs. Prior Year		1.9%	1.4%	1.6%	1.5%

The calendar year paid pure premium trend is between 1.4% and 2%, well below the "true" pure premium trend of 5%.

What would happen if an actuary did not account for the increasing exposure level distorting the trends in this example? If the actuary selects trends in line with what is produced by the calendar year data, then the selections will be too low. When trends are understated, then the indication will not be at an adequate level. If the company is not able to get to the appropriate rate level, the margins will be lower than needed and the price will be very competitive (if not too competitive). As a result, this might lead to a growth spurt with low margins.

2.3 Decreasing Exposure Level

The following chart contains the exposure for this example.

TABLE 10	
Calendar Year	Earned Exposures
X	100,000
X+1	100,000
X+2	100,000
X+3	90,900
X+4	78,500
X+5	63,475
X+6	48,575

Based on this exposure level:

TABLE 11					
	Calendar Year Paid Claims				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	4,000				
X+1	6,000	4,000			
X+2	10,000	6,000	4,000		
X+3		9,090	5,454	3,636	
X+4			7,850	4,710	3,140
X+5				6,348	3,809
X+6					4,858
All AY (total CY)	20,000	19,090	17,304	14,694	11,806
CY Pd Freq	0.2	0.2100	0.2204	0.2315	0.2430
Change vs. Prior Year		5.0%	5.0%	5.0%	5.0%

Similar to the increasing exposure level example, the use of calendar year data creates an artificial paid frequency trend. In this example, the paid frequency trend should be 0%, but the mismatch of paid claims and exposures has created a 5.0% trend.

The calendar year paid severity is:

TABLE 12					
	Calendar Year				
	Paid Losses				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	$1,600,000				
X+1	$1,200,000	$1,680,000			
X+2	$1,000,000	$1,260,000	$1,764,000		
X+3		$954,450	$1,202,607	$1,683,650	
X+4			$865,463	$1,090,483	$1,526,676
X+5				$734,802	$925,851
X+6					$590,432
All AY (total CY)	$3,800,000	$3,894,450	$3,832,070	$3,508,935	$3,042,959
CY Pd Severity	$190.00	$204.00	$221.46	$238.81	$257.75
Change vs. Prior Year		7.4%	8.6%	7.8%	7.9%

In this example, the paid severity trend is about 8%, above the actual 5%.

The calendar year paid pure premium is:

TABLE 13					
	Calendar Year Paid Losses				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	$1,600,000				
X+1	$1,200,000	$1,680,000			
X+2	$1,000,000	$1,260,000	$1,764,000		
X+3		$954,450	$1,202,607	$1,683,650	
X+4			$865,463	$1,090,483	$1,526,676
X+5				$734,802	$925,851
X+6					$590,432
All AY (total CY)	$3,800,000	$3,894,450	$3,832,070	$3,508,935	$3,042,959
CY Pd Pure Premium	$38.00	$42.84	$48.82	$55.28	$62.64
Change vs. Prior Year		12.7%	13.9%	13.2%	13.3%

Since the use of calendar year data overestimated the paid frequency trend and the paid severity trend, it is not surprising that the paid pure premium trend is overestimated. Additionally, when both trends are misestimated in the same direction the issue is magnified.

What would happen if an actuary did not account for the decreasing exposure level distorting the trends in this example? If the actuary selected trends in line with what is produced by the calendar year data, then the selections would be too high. If the trends are overstated, then the rate level indication will be higher than one produced from accurate trend projections. This may result in a price that is not competitive in the marketplace leading to a greater loss of business.

3 ADJUSTMENT TO CALENDAR YEAR DATA

Currently, actuaries have a few alternatives available to them.

The actuary can use reported claims instead of paid claims. The delay from the accident date to report date is shorter than the delay from accident date to close date. Since this time is shorter, the mismatch between claims and exposures is not as significant. However, there are a few drawbacks with using reported claims instead of paid claims. Using reported claims has the following disadvantages:

- Reported claims can be distorted by seasonality of reporting patterns. There could be spikes during different times of the year for things like claims office holiday schedules or a rush to file before the end of the year.
- Just because a claim is reported does not mean that it will ever be paid. For example, during periods of fraudulent activities there will be a significant increase in reported claims, but if these claims are found to be fraudulent, they will not translate into a paid claim.
- If there is an internal change in claim opening practice, the resulting numbers could distort the results.
- Using reported claims does nothing to address the problem with paid severity or paid pure premium.

Another alternative is to use accident year data instead of calendar year data. Accident year data will not have the problem of mismatching risk and exposures, nor will it have the same problem addressed above with the use of reported claims. The issue that accident year claim count data does have is that recent years are immature, so the data needs to be developed to ultimate. Loss development is a stochastic process, so there is inherent variability. As a result, there are multiple methods of loss development that could be appropriate to use. Since there is no established loss development method to be used in all situations, there is some subjectivity to using accident year data for loss trends.

The proposed solution is to attempt to match the risk with the appropriate exposure. The issue with calendar year data is that the paid claims in any calendar year may have come from older accident years, yet they are matched to the most recent calendar year exposures.

In the increasing exposure level example above, the number of paid claims in year X+6 was 25,763 and was matched to the 139,500 exposures. The reason that the paid frequency did not match the actual frequency is that the claims from accident year X+4 and X+5 had lower exposure level. Intuitively, it would make more sense to match these paid claims to the exposures that produced these claims. Using the notion from earlier in the article, the proposed formula is:

Adjusted Paid Frequency (APF) = $C_{0,12,X} / E_X + C_{12,24,X} / E_{X-1} + C_{24,36,X} / E_{X-2} + \ldots$

This formula can be thought of as adding the incremental frequencies. The incremental paid frequency in the first year ($C_{0,12,X}$ / E_X) is added to incremental paid frequency in the second year ($C_{12,24,X}$ / E_{X-1}), etc. The formula should end when all further $C_{T,T+12,X}$ are equal to 0.

The formula for pure premium is very similar to the one for adjusted paid frequency, replacing paid claims with paid losses:

Adjusted Paid Pure Premium (APPP) = $L_{0,12,X} / E_X + L_{12,24,X} / E_{X-1} + L_{24,36,X} / E_{X-2} + \ldots$

Since paid severity has to equal paid pure premium divided by paid frequency, the formula for adjusted paid severity (APS) is:

Adjusted Paid Severity = $(L_{0,12,X} / E_X + L_{12,24,X} / E_{X-1} + L_{24,36,X} / E_{X-2} + \ldots)/(APF)$
$= (L_{0,12,X} / E_X)/APF + (L_{12,24,X} / E_{X-1})/APF + \ldots$
$= ((L_{0,12,X} / C_{0,12,X}) * (C_{0,12,X} / E_X))/APF + ((L_{12,24,X} / C_{12,24,X}) * (C_{12,24,X} / E_{X-1}))/APF + \ldots$
$= (S_{0,12,X} * (C_{0,12,X} / E_X))/APF + (S_{12,24,X} * (C_{12,24,X} / E_{X-1}))/APF + \ldots$

The adjusted paid severity can be thought of as a weighted average of each 12-month accident year severity where the weight is the percentage that each 12-month segment contributes to the overall paid frequency.

3.1 Increasing Exposure Level using the Adjusted Formulas

The adjusted paid frequency is:

TABLE 14					
	Adjusted Paid Claim Frequency				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	4,000 / 100,000 = .04				
X+1	6,000 / 100,000 = .06	4,000 / 100,000 = .04			
X+2	10,000 / 100,000 = .10	6,000 / 100,000 = .06	0.04		
X+3		10,420 / 104,200 = .10	0.06	0.04	
X+4			0.10	0.06	0.04
X+5				0.10	0.06
X+6					0.10
Adjusted Paid Freq	.04 + .06 + .10 = .20	0.20	0.20	0.20	0.20
Change vs. Prior Year		0.0%	0.0%	0.0%	0.0%

When the adjusted paid frequency method is used, the paid frequency trend is 0% that matches what is assumed in the example.

The adjusted paid severity is:

TABLE 15					
	Paid Severities				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	$400[1]				
X+1	$200	$420[2]			
X+2	$100	$210	$441.00		
X+3		$105	$220.50	$463.05	
X+4			$110.25	$231.53	$486.20
X+5				$115.76	$243.10
X+6					$121.55
Adj Pd Severity	$190.00[3]	$199.50	$209.48	$219.95	$230.95
Change vs. Prior Year		5.0%	5.0%	5.0%	5.0%

[1] $400 = $1,600,000 / 20,000

[2] $420 = $1,680,000 /20,000

[3] $190.00 = $400 * .04 / .20 + $200 * .06 / .20 + $100 * .10 / .20

The use of adjusted paid severity formulas accounts for the mismatch of risk and exposures and accurately measures a 5% paid severity trend.

The adjusted paid pure premium is:

TABLE 16					
	Paid Losses				
Accident Year	X+2	X+3	X+4	X+5	X+6
X	$16.00[1]				
X+1	$12.00	$16.80[2]			
X+2	$10.00	$12.60	$17.64		
X+3		$10.50	$13.23	$18.52	
X+4			$11.03	$13.89	$19.45
X+5				$11.58	$14.59
X+6					$12.16
Adj Pd Pure Premium	$38.00[3]	$39.90	$41.90	$43.99	$46.19
Change vs. Prior Year		5.0%	5.0%	5.0%	5.0%

[1] $16.00 = $1,600,000 /100,000

[2] $16.80 = $1,680,000 /100,000

[3] $38.00 = $16.00 + $12.00 + $10.00

The adjusted paid pure premium formula measures the assumed 5% trend.

The adjusted formulas work under constant, increasing, or decreasing exposure level.

These formulas seem to work on a theoretical basis, but what about when actual data is used?

3.2 Actual Example

The data used in this example is hypothetical data from a personal lines insurance company.

The exposure level in this particular state was decreasing significantly:

TABLE 17	
Earned Exposures	
Calendar Year ending December 31	
Calendar Year	Earned Exposures
1998	60,249
1999	59,655
2000	53,760
2001	39,698
2002	21,525

In this example, the calendar year paid frequency is:

TABLE 18

Calendar Year Paid Frequency Trend

Bodily Injury Coverage

Date	actual data	6 pt. curve of best fit
9/01	3.97	4.107
12/01	4.61	4.575
3/02	5.23	5.096
6/02	5.79	5.677
9/02	6.44	6.325
12/02	6.78	7.046

REGRESSION	6 pt.
Avg Annual Trend =	54.02%

This is a significantly high paid frequency trend and should trigger some alarms. When the adjusted paid frequency formulas are used, the paid frequency is:

TABLE 19
Adjusted Paid Frequency Trend
Bodily Injury Coverage

Date	Actual Data	6 pt. curve of best fit
9/01	3.23	3.471
12/01	3.54	3.513
3/02	3.80	3.556
6/02	3.80	3.600
9/02	3.72	3.643
12/02	3.41	3.688

REGRESSION	6 pt.
Avg Annual Trend =	4.96%

The 5% trend is more reasonable than the 50+% trend that the calendar year data produced.

The calendar year paid severity is:

TABLE 20
Calendar Year Paid Severity Trend
Bodily Injury Coverage

Date	Actual Data	6 pt. curve of best fit
9/01	10,691	10,967
12/01	11,788	11,435
3/02	11,707	11,923
6/02	12,680	12,431
9/02	13,228	12,962
12/02	13,155	13,515

REGRESSION	6 pt.
Avg Annual Trend =	18.19%

In comparison, the adjusted paid severity trend is:

TABLE 21
Adjusted Paid Severity Trend
Bodily Injury Coverage

Date	Actual Data	6 pt. curve of best fit
9/01	10,228	10,597
12/01	11,194	10,782
3/02	10,800	10,971
6/02	11,436	11,163
9/02	11,654	11,358
12/02	11,144	11,557

REGRESSION	6 pt.
Avg Annual Trend =	7.18%

The 7% severity trend produced by the adjusted paid severity formula is closer to the inflation rate rather than the calendar year paid severity trend.

The calendar year paid pure premium trend is:

TABLE 22
Calendar Year Paid Pure Premium Trend
Bodily Injury Coverage

Date	actual data	6 pt. Curve of best fit
9/01	424	450
12/01	544	523
3/02	612	608
6/02	734	706
9/02	852	820
12/02	892	952

REGRESSION	6 pt.
Avg Annual Trend =	82.04%

The adjusted paid pure premium is:

TABLE 23
Adjusted Paid Pure Premium Trend
Bodily Injury Coverage

Date	actual data	6 pt. curve of best fit
9/01	330	368
12/01	397	379
3/02	410	390
6/02	434	402
9/02	434	414
12/02	380	426

REGRESSION	6 pt.
Avg Annual Trend =	12.51%

It is unlikely that the unadjusted calendar year paid pure premium trend can be thought of as being accurate, especially since it is known that the exposures are decreasing significantly.

4 CONCLUSION

This paper presents a theoretical solution that can be applied to real world issues. The method presented is not without drawbacks.

The premise of the method is to match risk to the exposure that produced the risk. Unfortunately, it is not practical to match every paid claim to the appropriate exposure, especially for long tail lines of business. In the hypothetical example in Section IV, most of the claims were paid within 8 years. Since most of the claims were paid within 8 years, all other paid claims from the 7th prior accident year or older were grouped together and matched to the 7th prior year's earned exposures. The 7th prior year's earned exposures were used since the exposure level then should be more reflective of the exposure level that produced the claims than the figures from the most recent year. In the example, since almost all claims are paid within 7 years, this is not a major drawback. It is outside the scope of this article to determine the optimal number of years to match the risk with exposure, because all years is not always a practical solution, but more than one is an improvement over current practices.

Another drawback of the proposed method is that it requires an extensive amount of data. For example, to calculate a 6-point annual calendar year paid frequency trend, an actuary needs 6 data points for 6 years of earned exposures and 6 data points for 6 years of calendar year paid claims. Under the proposed method, the actuary would need 13 data points for earned exposures and 48 data points for the 6 calendar year paid claims, with each year broken out by the most recent 8 accident years. Another weakness of this method is the method that data needs to be organized. The proposed method segments data into groups that are traditionally not used.

The other drawback of this method is the erratic results this method will produce when used for new lines of business. When companies enter lines of business, their exposure level will be low. Since this method matches claims/losses to exposures, there is a possibility that this method may produce results that are irrational. Although the adjusted formulas provide a more accurate result, credibility must be considered for small or volatile lines of business as with other methods of trending. Again, it is outside the scope of this article to determine the appropriate credibility standard for the results that this method will produce.

On the other hand, this method has multiple advantages. There is no need to assume a constant exposure level since risks are matched to the appropriate exposure. Also, there is no need to select development factors because calendar year data is still used. Finally, there is no need to make an assumption relating reported claims to paid claims.

The adjusted paid frequency, adjusted paid severity, and adjusted paid pure premium formulas are better alternatives to current practices since they eliminate the need to make major assumptions about the data and they provide a better match of risk and exposure.

Acknowledgments

I would like to thank Charles Parsons, Floyd Yager, Kathy Reither, members of the Committee on Ratemaking, and numerous others for their valuable assistance in reviewing earlier drafts.

Biography of Author

Chris Styrsky is an assistant actuary at Allstate Insurance Company in Northbrook, IL. His primary responsibility is personal lines pricing. He has a degree in Actuarial Science for the University of Illinois, Champaign-Urbana. He is a Fellow of the CAS and a Member of the American Academy of Actuaries.

Pitfalls in Evaluating Proposed Tort Reforms

Gail E. Tverberg, FCAS, MAAA

Abstract

Motivation. To provide the ratemaking actuary with a description of typical medical malpractice tort reforms and issues involved in pricing these reforms.
Method. The paper draws primarily on the author's own experience with reform legislation.
Results. Ten pitfalls in evaluating reforms are described.
Conclusions. For educating non-actuaries, the most important is Pitfall #8, relating to the difference between indemnity savings and the expected change to current rates. Actuaries pricing malpractice tort reforms should be aware of the issues underlying all ten pitfalls. Some of these issues may also be relevant to pricing reforms for other lines of insurance.
Availability. No additional material is available online.
Keywords. Ratemaking; medical malpractice; tort reform.

1. INTRODUCTION

Actuaries are frequently called upon to estimate the impact of proposed legislation. This is an area where differing techniques can produce very different estimates. In this paper, I will discuss some of the pitfalls I have encountered in evaluating medical malpractice tort reforms. Many of the issues involved would seem likely to apply to more general tort reforms as well.

1.1 Research Context

A search of actuarial literature shows only a two panels and one paper on related topics. At the 1987 Casualty Loss Reserve seminar, there was a panel on Tort Reform moderated by Jeffery Mayer, with Thomas Grillo and Phillip Miller as panelists[i]. Topics included claim file analyses and impacts of tort reforms on loss reserves.

There was another panel on Tort Reform at the 1989 Casualty Loss Reserve Seminar[ii]. This panel was chaired by Fred Kist, and included panelist speaking on the following topics:

Claus S. Metzner: Tort Reform Reserving Issues

Phillip D. Miller: Use of Individual Claim File Studies

Gail E. Tverberg: Impact of Tort Reforms on Loss Reserves: Lessons from Medical Malpractice

Material relating to these sessions is available on the Casualty Actuarial Society website.

Also, a paper by Allen Kerin and Jason Israel entitled, "The Analysis of the Effect of Tort Reform Legislation on Expected Liability Insurance Losses" is included in the 1998 Casualty Actuarial Forum, Including the Ratemaking Call Papers[iii]. This paper provides a proposed Insurance Services Office approach to evaluating the non-behavioral impacts on losses of a hypothetical general liability reform. The paper includes a discussion of the difficulties encountered when evaluating the impact of reforms.

1.2 Objective

The paper is intended to provide insights into the difficulties involved in evaluating tort reforms from an actuarial perspective.

1.3 Outline

In the "Background and Methodology" section, I will describe some of the more common types of reforms, for the reader not familiar with tort reforms. I will then discuss ten potential pitfalls in the "Results and Discussion" section.

2. BACKGROUND AND METHODOLOGY

2.1 Frequently Used Reforms

2.1.1 Caps on non-economic loss

Such provisions restrict the amount a plaintiff may recover for damages other than medical expenses, loss of wages and other direct economic costs associated with the injury.

2.1.2 Collateral source offsets

These reforms provide that the amount of an award will be reduced by recoveries from collateral sources, such as disability and medical insurance policies.

2.1.3 Limitations on joint-and-several liability

Under joint-and-several liability, each entity that contributes to an injury is individually liable for all or any part of the award. Thus, if one defendant has inadequate assets or policy limits to satisfy a judgment, a "deep-pocket" codefendant must pay a disproportionate share. Reforms limit the contribution of the deep-pocket defendant in certain situations.

2.1.4 Punitive damage restrictions

These reforms restrict the amount a plaintiff may recover as punitive damages to a multiple of the amount received for compensatory damages. In some instances, punitive damages are eliminated altogether or are limited to the most flagrant torts.

2.1.5 Periodic payments

For certain large awards, the portion attributable to future damages will be paid in regular installments over a fixed term or over the lifetime of the plaintiff.

2.1.6 Frivolous suit penalties

Typically, if a suit or a defense of a suit is found to be frivolous, the court may award attorneys' fees to the opposing party.

2.1.7 Limitations on attorneys' fees

In liability lawsuits, the plaintiff generally pays his attorney through a contingency fee, which is a percentage of the award or settlement. These reforms limit contingency fee percentages or provide for a court review of the reasonableness of fees.

2.1.8 Immunity statutes

Such statutes exempt certain individuals, institutions or public entities from tort liability under specified circumstances or place limits on the amount a plaintiff may recover from them.

2.1.9 Changes in pre-judgment interest

Many states provide that a jury award will be increased by a specified interest rate, between some specified date (such as the injury date or the date the suit is filed) and the date of the award. Recent reforms reduce the interest rate to correspond more closely to current interest rates.

2.1.10 Establishment of pre-trial hearing panels

Such reforms require that before a suit can be taken to trial, it must first be heard by a pre-trial panel. Under certain circumstances, the results of the panel may be admissible in court.

2.1.11 Establishment of state-operated funds to handle certain claims

States may establish a state-operated fund to cover claims in excess of a specified dollar amount, or to handle certain types of injuries, such as birth-injury claims.

2.1.12 Changes to the statute of limitation or statute of repose

Depending on the state, the statute of limitations can run from either the date of injury or from the date of discovery of the injury. If the statute of limitations runs from the date of discovery, there may also be a limitation on amount of time permitted for discovery, which is called the "statute of repose". Reforms can reduce the length of time permitted for either of these events.

2.1.13 Mandatory mediation

These require that mediation be tried before a case can go to jury trial.

2.2 Types of Reforms Discussion

This list of reforms is by no means exhaustive. State legislators are very creative in coming up with new types of reforms and combinations of reforms. Quite often packages of reforms will include items to try to appeal to a broad range of constituencies. Thus, besides including what one would usually consider tort reforms, such as the items above, there are often several other changes, as well, such as:

- Changes to the claim filing process, or to expert witness qualifications
- Requirements that insurance companies have rates approved in advance, perhaps after a mandatory hearing
- Mandatory rate reductions, if certain provisions are passed
- Requirements that physicians with adverse claims experience be investigated
- Various safety provisions, sometimes with requirements that physicians get premium reductions for making the changes

Thus, the legislation that is enacted generally includes quite a number of provisions, designed to appeal to different interest groups. Most often, the entire package of provisions needs to be reviewed, not simply a single provision.

2.3 Methodology

This paper draws primarily on my own experience in reviewing proposed malpractice tort reforms, working with closed claim data bases, and talking to lawyers involved with malpractice litigation.

3. RESULTS AND DISCUSSION

3.1 Pitfall #1: Not Adequately Understanding the Proposed Reform

Lawyers and others working with tort reforms frequently compile summaries, listing the basic provisions in the reforms package. While such a listing is helpful for getting an overview of the proposed legislation, it is better to obtain a copy of the legislation as proposed, and read it carefully. Pay close attention to details: If a non-economic damage cap is proposed, does it apply per defendant or for all defendants combined? Does it apply per plaintiff, or for all plaintiffs combined? If periodic payment of awards is proposed, how are payments treated after the death of the plaintiff? Do they stop, or does a portion of the payment continue?

Quite often there is also the issue of how the proposed legislation differs from current practice. As a starting point, one can look at the existing statutes, to see how they differ from what is proposed in the legislation. Additionally, it is also helpful to talk to someone who is familiar with current practices in the state (perhaps a lawyer handling claims) to get their view of how the proposed legislation differs from the current practice. There are situations where part of the existing statute has been declared unconstitutional, so it no longer applies, even though the wording of the statutes suggests the statute applies as originally enacted. There may also be situations where the existing statute is not enforced, raising questions whether the new statute will be enforced, either. A person familiar with existing practices can point out such pitfalls, and may also be able to provide insight as to how the current process really works and the reasons for the proposed changes.

3.2 Pitfall #2: Thinking Economic Loss Is a Uniquely Defined Amount

One of the more popular (and evidence suggests, effective) malpractice reforms is capping non-economic loss at a fixed amount, such as $250,000 or $500,000. With such a cap, the amount a claimant can receive in a jury award is limited to the claimant's economic loss, plus the cap amount. The question then becomes: How much is the plaintiff's economic loss?

If a person has only a minor injury, and is out of work a short time, the amount of economic loss is fairly easy to determine in retrospect. An insurer can ask how much the claimant paid in medical costs, and the amount of lost wages. There may be additional economic costs, such as the cost of transportation to medical treatments, to consider as well.

One somewhat tricky area, even on short duration claims, is the treatment of collateral source payments, such as health insurance and disability insurance payments. Depending on state law, economic loss will be either gross (full amount before recovery) or net of collateral source payments. If collateral source offset is permitted, the types of payments that are eligible for collateral source offset will vary by state. Besides privately purchased health and disability coverage, collateral sources could, at least theoretically, include payments by state disability programs, life insurance payments, Social Security disability income payments, Social Security health insurance payments, and Medicaid payments, among others. If collateral source offset is permitted, very often the premiums a claimant made to purchase the collateral source are considered as an offset to the calculation. If there is not an offset for collateral source payments, quite often the payer of the benefits will be allowed to subrogate against any award the claimant receives (collect back payments from the claimant after the claimant receives the award).

Another issue that sometimes arises is what the gross cost of hospital care really is. Hospitals charge different patients different amounts, depending upon who is paying for the care. In some jurisdictions, a plaintiff's attorney will be allowed to build his case based on the highest hospital rate anyone is charged, while in others, costs are limited to the actual cost for the particular claimant.

On longer duration claims, and on permanent injuries, evaluation of economic loss becomes more difficult. First, consider medical treatment. Will the claimant receive the finest medical treatment available, or treatment that is cost-effective and not quite state-of-the-art?

If the patient requires around-the-clock-care, will the patient be provided with private nurses in his home, day and night, or will care in a state-run hospital suffice? If the patient requires some extra care, but not around-the-clock nursing, can one expect that family members will provide the extra care at little cost, or must someone be hired (perhaps all day, every day) to assist in the care? If a family member provides care, will there be compensation for the wage loss of the family member? If the claimant is unable to walk, what kinds of additional help will be provided — a special car every few years, wheelchairs, changes to the claimant's home to accommodate a disabled person? The Virginia Birth-Related Neurological Injury Compensation Program, in its initial years of operation, provided each claimant with a specially-designed new home to accommodate the injury — a high cost way of dealing with the need for modifications to permit mobility.

If the person is disabled for a long time, or permanently, the question arises as to how long the medical care will continue. If the injured person is disabled, should one use a disabled lives table to determine life expectancy, or should a normal life expectancy be assumed? As an extreme example, a very injured infant may have a life expectancy of only a few years, but could live to be more than 70 years of age in the absence of the injury.

There is also the question of what inflation rate to assume on future medical costs. If payments are expected for many years, whether an inflation rate of 3%, 4%, or 5% is chosen (or something else) can make a substantial difference. If costs are to be net of collateral source offsets, one must also consider future collateral sources, and whether they will increase at the same rate as future medical costs. A person must also consider whether future costs should be discounted, and, if so, what interest rate should be used.

Wage loss becomes an issue on long duration claims as well. If the person was not in the workforce at the time of the injury, can one assume that in the absence of the injury, the person would have gotten a job? If so, what type of job, and at what wage? This issue applies especially to injured infants, since one would expect that the in the absence of the injury, the infant eventually would become employable. Wage loss also has the same issues of inflation and discounting as medical costs.

All of these issues make economic loss difficult to evaluate. In a typical suit, each side will present its own life-care plan for the plaintiff, and the expected future costs presented are likely to be quite different. If state caps on non-economic loss are enacted, proving economic damages becomes more important. Because of this, plaintiff's attorneys may apply more attention and creativity after caps become effective. Such changes may cause savings from the caps to be less than would be predicted based on the relationship between pre-reform loss payments and pre-reform economic losses.

3.3 Pitfall #3: Expecting Too Much of Closed Claim Data

Public databases, such as the National Practitioner Data Bank ("NPDB") and the Florida malpractice closed claim database, collect data from a variety of payers of malpractice claims: insurance companies, reinsurers, hospital self-insurance trusts, state joint underwriting associations, state excess funds, bankrupt insurers in runoff, and even individual physicians who do not purchase insurance. Trying to get everyone to report on a consistent basis is difficult, at best. For example, the NPDB summarizes claims by the year a closed claim report is received. The data at times appears quite "lumpy". Looking at the data, a person suspects that some organization has not been sending in reports, then sends in several years of back reports at one time. This kind of batching of claims can distort year- to-year comparisons and comparisons among states, especially if done on a single year basis. Averaging over a number of years tends to reduce this problem.

Economic and non-economic loss data collected through closed claim data has particular problems. One of the issues is the complexity of the calculation. If a claim has been settled prior to trial (and most are), there may not be enough information in the claim file to do a complete calculation of economic loss. Furthermore, as we saw in Pitfall #2, a person can get a very wide range of answers when calculating economic loss, depending on the assumptions used. If the completion of the closed claim form is assigned to a clerk, there is a significant possibility that shortcuts will be taken in completing the economic loss data. One of the more likely shortcuts is to include only past economic loss, and not try to estimate future economic loss. Another is to record a $0 or leave the field blank whenever the correct economic loss amount is difficult to determine. If either of these actions is taken, the amount of economic loss will be understated in the closed claim report.

Besides not having proper information to prepare economic cost estimates on claim settlements, there is also the issue that settlements reflect a variety of factors, besides the economic loss of the plaintiff. If the plaintiff has only a weak case, the settlement is likely to be lower than if liability is clear. It is not clear how one should apportion economic and non-economic loss, if liability issues are affecting the amount of the settlement.

When economic loss is collected on claims, one might think that the sum of the economic loss and non-economic loss recorded on a single claim (where there is only one payer) would add to the total amount of the claim. For the Florida closed claim data, this relationship rarely seems to hold (Exhibit 1). When we looked at the "current" Florida database, 1,304 out of 4,962 claims with indemnity payment showed neither economic nor non-economic loss, suggesting that no coding of economic or non-economic loss had been done. Another 1,774 out of the 4,962 claims showed some non-economic loss, but no economic loss. It would seem hard to believe that economic loss is absolutely $0 on these claims, since the average indemnity cost of these claims is $228,482, and most claimants have at least some economic loss. Of the 4,962 claims, 325 showed only economic loss. The economic loss on these claims amounted 1.85 times the claim amount. Only 1,255 out of the 4,972 claims showed both economic and non-economic losses. If an actuary uses this data to try to estimate the impact of caps on damages, he should be aware of its limitations. One approach might be to calculate a range of indications, with some of the indications excluding claims for which no economic loss is coded.

Very often, there are a number of different insurers or other entities making payments with respect to a single medical injury. This might happen when there are multiple defendants, such as a hospital and a number of physicians. It may also happen if there are layers covered by different organizations — for example, a self-insured hospital layer at the bottom, with a layer covered by insurance in the middle, and an excess insurer or an excess state fund on the top. In order to get a proper matching of indemnity with economic and non-economic loss, a person needs to be able to add together all the indemnity payments from various sources relating to a single medical injury and compare this amount to a single estimate of the economic loss. This is necessary since the plaintiff has only one economic injury, regardless of the number of organizations making payment of the claim. Getting everything matched together properly is tricky, and it is not clear that closed claim analyses can get all the pieces matched together properly.

Exhibit 1

Economic and Non-Economic Loss from Florida Closed Claim Database ("Current File")
Excluding Multiple Claims

Category	Count	Indemnity Payments	Reported Economic Loss	Reported Non Economic Loss	Ecomonic + Non Economic Loss Total	Average Indemnity Payment	Economic Loss/ Indemnity Payments
Total Excluding Multiple Claims	4,658	$1,103,066,068	$610,411,587	$623,733,883	$1,234,145,470	$236,811	55%
Econ > 0, Non Econ > 0	1,255	342,856,361	440,425,069	270,961,816	711,386,885	273,192	128%
Econ > 0, Non Econ = 0	325	91,842,261	169,986,518	0	169,986,518	282,592	185%
Econ = 0, Non Econ > 0	1,774	405,328,255	0	352,772,067	352,772,067	228,483	0%
Econ = 0, Non Econ = 0	1,304	263,039,191	0	0	0	201,717	0%

There can also be medical injuries with more than one claimant. Such medical injuries would typically involve a mother and infant. Matching these, as well as the multiple defendants and multiple insurers, adds a further level of complexity to closed claim data.

3.4 Pitfall #4: Misunderstanding the Phase-In Implications of the Reform

Reforms are likely to be effective on one of three bases:

1. Court awards after a certain date
2. Suits filed after a certain date
3. Claims arising from medical injuries after a certain date

The actuary needs to consider how the proposed effective date will co-ordinate with the type of malpractice coverage sold. If a reform is effective on claims arising from medical injuries after a certain date, this is equivalent to the reform being effective on an accident year basis. If the malpractice coverage is sold on a claims-made basis, the effect of such a reform will take several years to phase in, since the first year it will affect only claims with the accident year equal to the report year. Each year, a larger proportion of claims will be affected by the reform.

When a reform is effective with suits filed after a certain date, theoretically such a reform would match up fairly well with claims-made coverage, especially for an insurer that does not consider a claim to be reported until a suit is filed. The issue that one quite often encounters, however, is that lawyers have some discretion over when a suit is filed. Prior to the effective date of the reform, there may be a rush to file as many suits as possible. After the effective date of the reform, there may be relatively few suits filed for a while, since the "pipeline" has been emptied.

After a reform has been in effect for a while, it is possible that impact of the reform will change. This change could be in either direction. If the reform is one that trial lawyers can partially circumvent, it is possible that the reform will reduce in effectiveness over time. In the case of caps on non-economic loss, it sometimes appears that the reform reduces the annual trend increase in losses, and thus in some sense the reform tends to increase in effectiveness over time.

The constitutionality of many reforms is challenged in the courts. Until the constitutionality is upheld, a reform may not be fully effective. This seems to be the case with the set of reforms enacted in California (Medical Injury Compensation Reform Act or "MICRA") in 1975, but not upheld in the courts until 1985.

There have also been a number of instances where reforms were enacted (Texas, Illinois, Ohio, and Oregon, for example), and were at least partially effective. The reforms were later found to be unconstitutional, and malpractice costs in the state increased significantly after the constitutionality test.

3.5 Pitfall #5: Missing Indirect Impacts

Most reforms apply only to court awards, not to settlements. Clearly, if a reform results in lower awards, there will be at least some indirect impact on settlements. For example, if a state reduces the pre-judgment interest rate from 12% per year simple interest to 5% per year simple interest from the date a suit is filed to the date of the award, and average lag between suit and award is 4 years, the amount of prejudgment interest will be reduced from 48% of the award (on average) to 20% of the award (on average). The claims adjustor will consider the amount of the likely award (among other things), when making a settlement, so a change in award plus interest is likely to have an impact on settlements. It is not clear, however, that the percentage impact on settlements will be the same as on awards, since there are other factors affecting settlements. There may also be some other indirect impacts, such as a change in the willingness of the insurer to settle.

Other indirect impacts are less clear. For example, if a reform makes it much quicker and easier to pursue a claim, there are likely to be more people willing to pursue malpractice claims. For this reason, total malpractice costs may increase, even if the cost per claim is lower. Mandatory mediation is sometimes considered to have the potential to increase costs, because it makes settlements quicker and easier.

Changes to the compensation of lawyer through caps on contingency fees will increase the proportion of the award a plaintiff receives. It will also reduce the amount of payment the lawyer receives, and make him less willing to take on suits for which his compensation is likely to be too low. This may make the attorney less willing to take on suits for small dollar amounts, or suits where liability is unclear.

If a state establishes a state fund, and requires that health care providers purchase $250,000 limits from private insurers, private insurers may behave differently if they are providing $250,000 limits of coverage than if they are providing $1,000,000 limits (or more) of coverage. With low limits, the insurer may be less willing to defend the claim, and more willing to settle, since taking a suit to trial is likely to incur substantial attorney fees, and not save a large amount of indemnity payment.

3.6 Pitfall #6: Expecting a State Fund to Behave Like a Private Insurance Company

Quite often, states create state-operated excess funds to provide coverage in excess of some required primary limit. Florida and Virginia have also established state-sponsored birth injury funds. On paper, these funds look much like private insurers, but in practice they often behave quite differently. Some of the differences include:

- Adequacy of Funding. Some state funds are pay-as-you-go by design. Others are intended to be fully funded, but may not be as fully funded as an insurance company, because of the pressures to keep rates low, and because of the uncertainties of funding for excess limits.

- Expenses and other costs. Expenses of state funds may be significantly lower than for an insurer, particularly if there is no commission expense. Also, state funds generally do not pay federal income tax. Funds may not to have the equivalent of shareholders' equity, and even if they do, are not likely to be concerned about making an adequate return on equity.

- Willingness to settle. Without a profit motive, some state funds may be more willing to settle than a private insurer would be. The willingness to settle may also be related to a desire to help the injured person.

- Background of employees. If wages are constrained by state budgets, the employees may be less experienced than their counterparts at private insurers.

3.7 Pitfall #7: Missing Additional Expense Resulting from the Reform

There are several types of reforms that are likely to results in extra cost because of additional administrative expense and other factors. It is possible the same reforms will result in some savings as well, but the actuary will want to consider both when estimating the net savings or costs. Some reforms that are likely to result in addition expense are the following:

- Pre-Trial Hearing Panels. In many cases, the insurance company will need to present its case twice: First to the pre-trial hearing panel, and second at the trial itself. Thus, there may be additional legal expense because of the additional work involved. If there are no incentives built into the program to hear cases quickly, the pre-trial hearing panels may significantly delay the regular trial. If the jury trials are delayed, there could be other additional costs as well—higher indemnity costs because of the delay, and possibly issues of witnesses no longer being available, because of the delay.

- Patient Compensation Funds (PCFs). PCFs are state-operated funds that provide excess coverage over required underlying insurance, purchased in the regular marketplace. Typically, the required underlying coverage is in the $200,000 to $1,000,000 range. One issue that arises is the additional administrative cost of the PCF. In the absence of the PCF, the underlying insurer would write higher limits, so that only one insurer would be needed. In these cases, having the PCF means there will be an additional set of expenses arises because both entities will have their own administration and claim-related costs, even if they are not entirely duplicative In addition, these entities will need to prepare their own financial statements, and will need to made their own investment decisions.

- Higher policy limits. Any change that causes physicians to purchase higher policy limits is likely to result in higher costs. For example, a cap on total damages that is higher than physicians the policy limits physicians traditionally purchase can result in physicians purchasing higher limits, either by choice, or as a result of a requirement of a hospital in which the physician practices. Higher policy limits mean higher costs for two reasons:

— Because of the increased limits factor differential an actuary would expect.

— Because of a change in the economics of the situation. If policy limits are raised for a significant share of insureds, the change in policy limit may change the claim environment as to what is an appropriate settlement or award, and may result in higher payments to claimants for the same injury. This impact is the reverse of the capping effect of policy limits that sometimes occurs if policy limits remain at a relatively low level, in comparison to awards.

- Added features to make the program more balanced. Very often, if a reform gives some benefit to insurers, there will be other provisions added to make the bill more balanced. For example, malpractice insurers may be required to submit data for a closed claim study, or malpractice insurers may be required to obtain prior approval for any rate increase. In one proposal, malpractice insurers were required to notify insureds of any potential rate increase, and to have a hearing and approval before any rate increase could be implemented. These changes could make it difficult to collect adequate premium for the coverage sold.

3.8 Pitfall #8: Forgetting the Difference between Anticipated Indemnity Savings and Expected Change to Current Rates

An actuary will want to look specifically at what types of changes are expected: indemnity, legal expense, or both. For many types of reforms, including caps on non-economic loss, requirements for periodic payments, and changes in pre-judgment interest, the change will be predominantly affect indemnity costs.

The actuary will want to consider how the change to indemnity costs can be expected to affect overall costs. Suppose indemnity costs before a tort reform are $100, and are expected to be $90 after the tort reform. Suppose legal costs are $30, both before and after the reform, and suppose other costs are $20, both before and after the reform. Total costs are then $150 before the reform, and decrease to $140 after the reform. The percentage savings are then $10 / $150 or 6.7%, rather than the 10% some non-actuaries might expect.

A related issue is the adequacy of the current rates. If rates are quite inadequate, prior to the tort reform, the tort reform may bring the rates closer to an adequate level. The actuary will want to consider the impact of the tort reform in determining the appropriate rate change. In some cases, a rate increase may still be needed, even with the reform. This idea is easy for an actuary to see, but may not be as obvious to legislators.

3.9 Pitfall #9: Failing to Consider Differing Impacts by Policy Limit and by Direct Insurer vs. Reinsurer

A wide range of medical providers purchase professional liability insurance. Besides physicians and surgeons, dentists, chiropractors, nurses, optometrists, physical therapists, and many other health care workers purchase malpractice coverage. Many types of institutions including hospitals, nursing homes, assisted living centers, clinics, and surgical centers also purchase malpractice coverage. Each of these types of providers has a different mix of claims, with the average size of claim varying with the type of provider. For example, dentists and physical therapists typically have quite small claims, while physicians and hospitals have larger claims.

Policy limits also vary greatly. In a few states, policy limits as low as $200,000 to $500,000 per claim are common for physicians. In other states, $1,000,000 limits are common for physicians. Hospitals, nursing homes groups, and others that have significant assets to protect very often have much higher policy limits, as much as $20,000,000 or more per claim.

When legislators enact a package of reforms that is expected to have a significant impact, such as a cap on non-economic loss, legislators may consider requiring insurers to reduce malpractice rates by a selected percentage, such as 10% or 20%, to reflect the expected cost savings of the reforms. Because of the diversity in types of malpractice coverage sold and policy limits, insurers are likely to be impacted differently by a proposed cap, so this flat percentage reduction is not very equitable. For example, a $500,000 cap on non-economic loss is likely to provide much more benefit to a hospital with a $20,000,000 policy limit than to a physician with a $200,000 policy limit, since the low policy limit already provides some capping effect on claims.

A related issue is the impact of reinsurance. An insurer that writes coverage on a direct basis will be the one affected by a mandatory rate reduction. A reinsurer is free to charge whatever rate it chooses. In the case of caps on non-economic damages and other reforms affecting large claims, the most significant benefit will be with respect to layers which are typically reinsured. Therefore, a rate rollback, even if theoretically correct in total for a direct insurer and its reinsurer, may result in too little money for the direct writer of the coverage if the reinsurer is not willing to reduce its rates.

3.10 Pitfall #10. Special Considerations for One Line / One State Insurers and Self-Insurers

When a state passes tort reform legislation, the exact amount of the benefit is not clear in advance. There is also often a question of whether the reform will be upheld in the courts, as mentioned in Pitfall #4. If a reform is found to be unconstitutional several years after it is passed, an insurer may find it may have to pay more indemnity than planned on several past coverage years. The insurer cannot retroactively raise rates, and will need to cover any losses from such a change with its surplus.

A number of medical malpractice insurers are provider-owned insurers, writing coverage primarily (or entirely) in one state. The question arises: How optimistic should these insurers be in reflecting the expected benefit of the tort reform changes in their rates? These companies typically pay dividends to their policyholders, so have the option of returning extra premium later, if it is not needed. Since the company is provider-owned, any overcharge will remain in the company, and will be owned by the physicians (or others) who paid a higher-than-necessary premium. These companies were formed by physicians (or other providers), for the purpose of providing coverage to their members, because other coverage was not readily available. Continuing to provide a market for malpractice insurance is thus one of the primary purposes of these companies. If the company should overestimate the impact of the reforms, or have the reforms overturned in the courts, it could find itself in serious financial difficulty, since it cannot spread its risk to other lines or other states.

Given these consideration, one-line, mostly one-state companies often choose to be cautious in recognizing the benefit of tort reform. Because of competition, these insurers cannot ignore the impact of tort reform. But given the serious difficulties that could result if the rates are set too low, and the possibility of returning funds later through dividends, taking a cautious approach may seem best. Otherwise, these companies may find themselves in financial difficulty and because of this, cease to provide the malpractice market for which they were established. The physician-owners would not find this an acceptable outcome.

4. CONCLUSIONS

Because the many potential pitfalls in evaluating tort reforms, the actuary will want to evaluate any proposed reform carefully, and consider the many issues involved. Actuaries who evaluate the expected costs impact of proposed malpractice tort reforms should be aware of the issues underlying all ten pitfalls. Some of these issues may also be relevant to pricing reforms for other lines of insurance.

Of the various pitfalls discussed, probably the most important from the point of view of the non-actuary is the difference between expected indemnity savings with the expected change to current rates (Pitfall #8). If the legislature of a state enacts a reform that is expected to result in a 10% reduction in indemnity, it is easy for a legislator to jump to the conclusion that rates should be reduced by 10%. The actuary needs to be careful to explain the various issues involved, including the importance of the adequacy of current rate level, and the need to adjust for other components of rate level.

In this paper, I have discussed only pitfalls closely related to actuarial analyses of tort reforms. There are closely related areas, each with their own pitfalls. For example, there are a number of studies relating to the frequency of iatrogenic (caused by medical practice) injuries and iatrogenic injuries caused by medical negligence. Looking at the pitfalls of these studies is outside the scope of this paper.

5. REFERENCES

[i] 1987 Casualty Loss Reserve Seminar, Session 1D, Tort Reform, 104-137.

[ii] 1989 Casualty Loss Reserve Seminar, Session 2E, Tort Reform, 299-339.

[iii] Kerin, Allan and Jason Israel, "The Analysis of the Effect of Tort Reform Legislation on Expected Liability Insurance Losses," Casualty Actuarial Forum, Including the Ratemaking Call Papers, Winter 1998, 153-192.

Abbreviations and notations
PCF, patient compensation fund

Biography of the Author

Gail Tverberg is a consulting actuary for the Tillinghast business of Towers Perrin. She is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries. She has a B.A. in mathematics from St. Olaf College in Northfield, Minnesota, and an M.S. in mathematics from the University of Illinois. Ms. Tverberg's primary area of expertise is medical malpractice. She is located in Atlanta, Georgia.

Prior to joining Tillinghast in 1981, Ms. Tverberg worked for CNA in Chicago.

Insurance Industry Decision Support: Data Marts, OLAP and Predictive Analytics

George Bukhbinder, Michael Krumenaker, and Abraham Phillips

Abstract

Motivation. Data Warehouses and Data Marts increase the power and efficiency of an Insurance company's Business Intelligence capabilities by supporting queries, OLAP and data mining. Web-enabling of these applications makes them more user-friendly. The potential benefits greatly outweigh the costs. Data warehouse/data mart implementation streamlines information delivery for decision support and significantly simplifies development of general linear predictive models that have become more popular with actuaries and statisticians in recent years.

Method. A data mart and OLAP system was implemented for a major property and casualty insurance company.

Results. Successful implementation substantially improved the insurer's operational efficiency, providing enhanced analytical and reporting capabilities.

Conclusion. Business needs and business strategy must drive decisions about the structure and functionality of the Business Intelligence platform including the data warehouse and the data mart. The data warehouse must be well planned for the organization to realize the expected efficiencies and process improvement. However, even with a well-designed data warehouse and up-to-date software tools for data access and analysis, it is critical that the enterprise build and maintain its analytical expertise. Therefore, actuaries play a critical role in maximizing of benefits that these new tools can offer.

Keywords. Business intelligence, OLAP, data warehouse, data mart.

1. INTRODUCTION

This paper discusses certain important principles and issues in the development and operation of data warehouses from a business perspective, with a special focus on data marts, including OLAP and predictive modeling capabilities. The case study will be an actual system (the "P&C Data Mart") created for a major property and casualty insurer.

Business Intelligence, the process of gathering, storing and analyzing data, building knowledge from the analysis and taking action based on the knowledge is the single most powerful success factor in business today. In an insurance company, nobody understands data and its value better than the actuary. An actuary is in the best position to provide visionary leadership to the organization's efforts to develop analytical capability and expertise.

Consolidation within the industry, mergers, acquisitions and divestitures involving insurance and other financial services providers have created a challenging business environment. Technological advances are causing major changes in the insurance sales and distribution system. Insurance companies are recognizing the need for early detection of changes in the environment and quick responses to those changes. In this volatile environment, competitive comparisons and data analyses need to extend beyond pricing and actuarial applications. Marketing, customer retention, sales force management, underwriting selection, pricing, claims fraud detection, loss reserving, risk management and all other aspects of the insurance business could benefit by using Business Intelligence tools.

Successful innovators in the insurance industry have been improving their Business Intelligence capabilities over the years. They have been building data warehouses and data marts. They have been using tools for on-line analytical processing (OLAP) and predictive modeling (data mining) to convert raw data into strategic advantage. They are now reaping the benefits and building on their successes.

Combining automated preparation of transaction data into account-level and more highly summarized tables for inclusion in the data mart with user-friendly means of accessing the information over the Internet or the enterprise Intranet allows the analysts to focus on analysis and research.

2. DATA WAREHOUSES, DATA MARTS, OLAP, PREDICITVE ANALYSIS

2.1 What to Expect from a Data Warehouse or Data Mart

A data warehouse is the foundation of powerful analyses. It supports business decision-making by allowing managers and analysts to examine data and perform powerful analysis easily and quickly. It facilitates measurement of the effects of various combinations of factors (geographic, demographic, rating and underwriting variables) on sales, premium, losses, loss frequency, loss severity, loss ratio, customer retention and other measures, and provides a strong platform for regression analysis and various other forms of predictive analysis.

Data warehouses, OLAP and data mining tools will not, by themselves, make a company successful. Data warehouse development must be driven by clearly understood business needs. First, the company must understand its business needs and what factors are important to its success. Then it must develop and implement ways to meet those needs. An actuary with a broad strategic vision extending beyond the confines of ratemaking and reserves is an ideal person to lead an enterprise in developing its analytical capabilities. The ability to anticipate future needs plays a key role in the success or failure of a data warehouse.

2.2 The data warehouse

The term "data warehouse" is often used in different contexts to mean different things. In this section, we discuss three types: Corporate Data Warehouses, Data Marts, and Operational Data Stores.

Ideally, a company would like to have a "single version of the truth" in one large Corporate Data Warehouse so that all data used for reporting and analysis is extracted from it. Such a data warehouse will contain a large amount of detailed transaction-level historical data that covers multiple subject business areas brought together from multiple sources, and integrated into a convenient format for extracting information for building Data marts for individual departments and for other uses that require detailed, granular historical data. In practice, a large company may have more than one data warehouse, but not too many.

Data marts are built to address the analytical needs of individual departments. For example, while Actuarial and Underwriting areas could possibly share a data mart, Marketing may need to have a separate data mart oriented to its specific needs and the Claims Department may have to have still another data mart. Like the larger data warehouse, data marts typically contain historical data. Selected data is summarized to a level adequate to meet the intended analytical needs, for inclusion in the data mart. For example, actuaries typically do not need many items of data that might be of interest to Claims professionals. The data for the data mart may come either exclusively from a data warehouse or certain operational systems, or both.

Many experts advise against building data marts before completing an enterprise-wide data warehouse. They also prefer to have all the data for the data mart come from the data warehouse. They fear that otherwise, data cleansing efforts will be inadequate and proliferation of independent "stove pipe" data marts will result in many inconsistent "versions of the truth", resulting in indecision and frequent and expensive efforts at reconciling data sources.

Operational Data Store, unlike the data warehouse or the data mart, contains near-real time data captured from operational systems. This data is used for tactical analysis to support on-line operations.

Data warehousing is an on-going process, rather than a "once and done" effort. As the company and the business change, the data warehouse, operational data stores and data marts need to evolve with them. New data will have to be captured, and analytical tools have to be developed and continuously improved.

2.2.1 Transactional data vs. the data warehouse

Actuaries across the industry have been using summarized, historical data bases for their analyses. However, until recently the same may not have been the case for other areas such as marketing, underwriting and claims. Even the databases used by actuaries have been limited in scope. Typically, extracting data from transactional databases has been necessary for many analyses. This approach has several drawbacks:

1. Extracting information from transactional tables can be immensely complicated.
2. Processing time would often be prohibitive, since transaction tables may contain millions or tens of millions of lengthy records.
3. Data validation and data cleansing are virtually impossible to perform while extracting data from transactional data bases for immediate analysis.

A data warehouse or data mart will contain data that has already been validated, cleansed and preprocessed at the beginning of each processing cycle, which might be weekly or monthly. This data extraction and summarization makes the data suitable for analytical queries (e.g. "What is the distribution of all of our sales by state and line of business?"). By assembling and preprocessing data we avoid having to perform the same resource-intensive steps again and again throughout the processing cycle.

2.2.2 The case for building a data warehouse

Business needs and business strategy must be the driving forces behind the building of a data warehouse. If the data warehouse does not meet compelling business needs, then don't invest the time and money in building it.

How do you make a business case for building a data warehouse? A Business Intelligence effort can be successful only in a company with visionary leadership. If the purpose of building a data warehouse is simply to automate existing processes, it may be difficult to find enough value to justify the undertaking. Automation may simply make an inefficient process run faster. The resulting savings may be minimal. On the other hand, when innovation and business process re-design becomes the goal, data warehousing becomes part of a real Business Intelligence effort that would ultimately deliver significant competitive advantage.

The person championing a Business Intelligence project proposal will need to understand its basic precepts and believe in them. This person needs to be an excellent communicator and be willing to take calculated risks.

Trying to do something simply because "our competitors are doing it" is an approach fraught with danger. Movements such as Total Quality Management (TQM) and Re-engineering have amply demonstrated this. Every company is unique with regard to its culture, employees, customers, business processes, organizational structure and many other factors.

If budget is tight, the only option may be to start small. Build a data mart for a specific application that might, in a relatively short time, generate a high ROI from expense savings or from process redesign and efficiency improvement. Success breeds success. Once you establish your credibility, making the business case for a larger project becomes much easier.

Nonetheless, making the business case always involves identifying the business needs and the company's "points of pain" in current processes. Success in making the business case depends on how the issues are presented, the conviction and initiative demonstrated by the "champion", the credibility of the "champion" and the strategic perspectives of the "champion" and the senior executives of the company.

Some of the areas where business needs or points of pain exist may include the following:

– responding quickly to business strategies of more nimble competitors
– effectively coordinating pricing strategies and underwriting rules and activities
– performing innovative actuarial analyses (e.g. exploration of new discounts, rules-based pricing, and data mining)
– performing reliable periodic analyses of the automobile class plan, deductibles and increased limit factors, taking into account the impact of other rating variables and a changing business profile
– responding on a timely basis to regulatory agency inquiries and rate filing requirements
– understanding the causes for declining customer retention

- putting together rate filings fast enough to keep pace with competitors
- efficiently performing sophisticated analysis by state (e.g. making use of California Sequential Analysis)

A business case document should provide specific details about each need or the consequences of each problem. It should also describe how the data warehouse and related tools (such as an OLAP system and data mining tools) will enable the enterprise to address each of the issues, and how doing so could provide the enterprise with a competitive advantage.

Practically speaking, the business case will be most effective if the project's "champion" is willing to stick his or her neck out and take responsibility for demonstrating the business advantages to be gained once the warehouse and associated tools are in place.

2.3 OLAP

As with "data warehouse" and "data mart", there seems to be no single agreed-upon definition of OLAP, but a reasonably good one is "[a] category of applications and technologies for collecting, managing, processing and presenting multidimensional data for analysis and management purposes."

OLAP takes the analyst beyond pre-defined reports and allows him the freedom to delve deeper in directions suggested by the data, searching for trends and anomalies.

2.3.1 Analysis Variables (Facts) and Class Variables (Dimensions)

Analysis variables or facts are the quantities being measured. Analysis variables in the P&C Data Mart include premium earned, the number of claims, the amount of loss incurred and paid, and allocated expenses. Class variables or dimensions are variables whose effects are being measured individually or in concert with other class variables or dimensions. For auto insurance, class variables include the type of coverage (e.g. BI, PD, Comprehensive, Collision), geography (region, state and territory), coverage limits, deductibles, number of times renewed, safe driver insurance points, driver training, age or years of driving experience, and time – month, quarter, or year.

2.3.2 An OLAP Example

In an Excel spreadsheet, a user can view the effect of two class variables at the same time in a matrix format. But modern OLAP technology allows simultaneous analysis by additional dimensions with minimal effort. In Figure 1, the measurements are earned premium and incurred loss. The dimensions selected were peril and state in rows vs. use and time in columns.

Figure 1 contains the first eight columns and all the rows in the initial screen, showing measures of analysis variables (earned premium and incurred loss) for Peril by Use. If the user wants to look at the distribution of earned premium and loss among all states for a particular type of coverage, e.g., accidental death and dismemberment, he or she simply opens up the ADD row, as in figure 2. This opening is referred to as "drilling down" through one dimension into the next. Similarly, by clicking on any "use" category, data for that use category for the various quarters could be displayed. Thus, the analyst is able to look at data from a variety of angles and continue to explore reasons for any special or abnormal results.

Without multidimensional views available on demand, analysts must draw information from various tables in the data mart or other sources and perform analysis either through user-oriented tools such as electronic spreadsheets, or performing complex database queries - a skill not every analyst has the time or desire to acquire. Multidimensional views such as figures 1 and 2 can be exported to Excel spreadsheets or other databases for independent analysis or display. Fast response and the ability of non-programming professionals to perform analysis with no programmer involvement is a clear advantage of OLAP. The advent of OLAP essentially makes standard spreadsheet reporting obsolete.

2.3.3 Multidimensional Databases (MDDBs) vs. Star schema

There are two competing OLAP methodologies based on the data structure used: multidimensional databases and star schemas.

Transactional data are typically captured in relational databases that store data efficiently in tables that are linked by primary key – foreign key configurations. Such relational database systems, designed for transactional data, were unsuitable for data analysis. This led to the development of a modified form of relational data structure called star schemas. A typical star schema is pictorially represented with a "fact" table that holds measurements (e.g. premium, losses) in the center, surrounded by tables for "dimensions" (e.g. annual mileage, use, coverage limits). Star schemas offered substantial improvement in performance in

	Use	Quarter								
		⊞ Bus/WL		⊞ Business		⊞ Farm		⊞ OTHER		⊞
Peril	State	Earned Premium	Incurred Loss	Earned Premium	Incurred Loss	Earned Premium	Incurred Loss	Earned Premium	Incurred Loss	E
⊞ ADD		$186	$	$156	$					$
⊞ APIP		$2,613	$727	$5,231	$116,740	$378	$			$
⊞ BI		$728,084	$470,816	$739,500	$34,865	$39,153	$193,646	$399,896	$342,161	$
⊞ COLL		$553,351	$273,222	$581,352	$163,662	$15,217	$42,406	$158,170	$183,222	$
⊞ COMP		$283,799	$190,524	$307,895	$61,206	$11,198	$3,869	$71,191	$14,005	$
⊞ MED		$45,461	$45,555	$37,616	$2,237	$3,157	$	$727	$	$
⊞ OBEL		$312	$	$380	$	$31	$			$
⊞ PD		$371,138	$238,722	$369,020	$93,056	$20,790	$28,757	$216,363	$121,572	$
⊞ PIP		$185,230	$48,233	$207,175	$110,220	$11,328	$3,527	$26,074	$24,289	$
⊞ PPI		$940	$	$531	$	$56	$	$22	$	$
⊞ TOWING		$4,029	$967	$6,308	$1,671	$318	$125	$453	$497	$
⊞ UIMBI		$	$							$
⊞ UMBI		$150,121	$2,470	$205,658	$1,550	$20,201	$	$55,290	$	$
⊞ UMPD		$236	$	$146	$			$	$3,973	$
Grand Total		$2,325,500	$1,271,236	$2,460,968	$585,207	$121,827	$272,330	$928,186	$689,719	$

Figure 1: Initial OLAP Screen

Use	Quarter						
		⊞ Bus/WL		⊞ Business		⊞ Farm	
Peril	State	Earned Premium	Incurred Loss	Earned Premium	Incurred Loss	Earned Premium	Incurred Loss
⊟ ADD	AL						
	AR						
	AZ	$16	$	$20	$		
	CA			$20	$		
	DE						
	FL			$16	$		
	IA	$16	$				
	ID						
	IL	$36	$				
	IN	$20	$	$40	$		
	KY	$12	$				
	LA						
	MD	$16	$				
	ME						
	MN						
	MO						
	MS						
	MT						
	ND						
	NE						
	NM						
	NV						

Figure 2: Drill-down into State dimension

analysis situations, over traditional relational data structures. In the star schema structure, dimensional relationships are not reflected in the way data is stored physically; they need to be created from multiple tables through primary-foreign key relationships.

On the other hand, data in multidimensional databases are organized by preserving the multidimensional relationship. There are no primary key-foreign key references needed. Each fact or measure is stored as a value indexed by the values of the dimensions. The data is thus stored in a simple format and could be retrieved faster, but takes up more storage space.

In multi-dimensional OLAP, data is pre-summarized into "cubes" that are intuitively descriptive of an n-dimensional structure, the number of dimensions being equal to the number of class variables (e.g. State, age, marital status). The structure also accommodates drill-down through unrelated (age→state) or increasingly granular (region →state → territory) hierarchical structures.

Relational OLAP vendors using star schemas have made their products competitive by optimizing their performance, but for more complex OLAP queries, particularly of those involving complex calculations, the multidimensional approach may still hold an advantage.

2.3.4 Four Standards for a Successful OLAP system

Timeliness:	Requests must be processed reasonably quickly – in seconds or minutes, not hours or days
Understandability:	Reports must be self-explanatory.
Ease of use:	Analysts must be able to create reports easily, without programming
Access:	The system must be easily accessible from different locations

2.4 Predictive Analytics (Data Mining)

MDDBs are intended for exploratory analysis. More sophisticated analytical tools and models are required to derive actionable results. Unfortunately, some vendors present MDDBs as tools for data mining. They also tend to discount the need for analytical expertise.

Of course, we could develop and set up routine processes for fraud detection and other operational uses. These processes are often designed using rules derived from extensive research and data mining. However, running the routine processes in itself is not data mining, but simply an application of the results of predictive analytics.

"Predictive analytics" is becoming the preferred term over "data mining". The convergence of technology, mathematical statistics, probability and other disciplines has resulted in highly powerful techniques for data analysis and prediction. Farmers Insurance's success with data mining, particularly the identification of the market segment of sports car owners who also own a typical family car, has become almost legendary.

In recent years, predictive modeling using general linear models (e.g. Poisson regression, logistic regression, log-linear analysis) have become immensely popular among actuaries and statisticians. Such modeling has the advantage of being more tractable and more amenable to meaningful interpretation than results from neural networks and classification tree analysis. Highly sophisticated software such as the IBM Intelligent Miner and SAS Enterprise Miner as well as many specialized software products with more limited functionality have put data mining within the reach of analysts who are not necessarily expert statisticians.

Predictive analytics can be one of the most critical uses of the data warehouse. Skillful analysis of customer data can address analytical challenges such as

- identifying new pricing variables
- finding and accounting for the overlap and interaction of underwriting and pricing variables
- assessing the impact of rate change on customer retention by market segment
- profiling and clustering current clients and prospects
- profitability models at different levels (national, state, agent), retention, cross-selling, renewal underwriting, new business acquisition
- separate models for different segments, such as tort vs. no-fault states, or "special" states (e.g., New Jersey) vs. other states
- different models by coverage (e.g. personal injury vs. property damage)

- alternatives to current insurance scoring based on variables used in credit scoring
- prospective vs. retrospective (claim vs. no-claim) models and sampling

A data warehouse can support such mission-critical business objectives effectively only if it is designed to do so. The best application software provides no benefit if the available data is inadequate or the data structure does not facilitate efficient analysis. MDDBs and star schemas[1] are not the best data structures for predictive analysis. The data structure must support easy access and data manipulation by the analytical software.

3. IMPLEMENTATION

Our discussion of implementation will consist of three sections:

1. General considerations
2. The case study: actual implementation of a Data Mart.
3. Reducing the need for re-programming when business rules change (an opportunity for future cost savings).

3.1 General considerations

We cannot overemphasize the fact that business needs are most critical in determining the contents and capabilities of the Business Intelligence system. Often such projects fail because they are built on the assumption that "if we build, they will come". Even having the right software and top management support may not provide the expected benefits unless there is buy-in from the users.

3.1.1 The actuary as visionary and essential element in implementation

The process of building the data warehouse should go hand in hand with building analytical expertise. That is what makes the visionary role of the actuary critical. Having world-class OLAP tools to "slice and dice" and "drill-down and drill-up" and access to the best data mining software for clustering and neural networks will yield little benefit if the analysts do not continually build, upgrade, refine and refresh their skills. Consultants and vendors do a disservice to their clients if they tout tools over expertise. While such tools will enhance the analysts' ability to choose analytical methods, compare results derived by different methods and interpret results, ultimately enabling purposeful action, they are never a substitute for analytical expertise.

3.1.2 Data Warehouse requirements

Ideally, we would have one huge data warehouse that holds all the data from internal and external sources that would be shared across the enterprise. But business needs, data, reports, applications and access requirements are so diverse that the challenge of designing and building an enterprise-wide data warehouse is too formidable.

The decision about data warehouse contents may benefit from

- a review of past analyses by the enterprise
- a literature review of analyses by academics and professionals in related fields
- a review of the business environment and challenges that lie ahead
- a review of analyses that the warehouse could support

Once the content has been selected, we must identify the data sources. These could include the following internal data sources:

- existing data warehouses and data marts
- billing systems
- transactional tables
- a POS database

In addition, there are external sources:

- Choice Point
- CLUE
- Current Carrier
- ACXIOM
- Census
- RL Polk
- Regulatory information sources

- Weather data by zip/county (RMS, AIR, Guy Carpenter)
- Competitors' rate filings and territory definitions
- Other competitive information

How the data is structured within the data warehouse is determined by the intended uses of the data. The data warehouse/data mart should be carefully designed to support the desired analytical techniques, e.g., queries, OLAP, cluster analysis, regression, decision trees, and neural networks. Therefore, business users should take a very active part in the requirements gathering phase.

3.1.3 Project Management

Building a data warehouse is best accomplished through a combination of traditional Systems Life Cycle Methodology and an iterative process incorporating prototyping and double-loop learning. Although a rigid Systems Life Cycle approach helps create discipline and project stability, there will be many change requests along the way so that change management efforts could create more confusion than discipline. Double-loop learning involves thoughtful adjustment to strategies, conceptual frameworks and action plans, based on problems and issues identified. A certain amount of judiciously managed prototyping and double-loop learning will enhance the flexibility and quality of the Warehouse.

Planning the data warehouse/ data mart must include

- becoming thoroughly familiar with the data sources and the data in them
- evaluating the available software (e.g. UDB, Oracle, Sybase, SAS, RedBrick)

Business and analytical needs drive the choices of content, functionality and user-friendliness, but the data warehouse developers are responsible for prescribing the data types, table structures, methods of data extraction, data cleansing, transformation and loading, information delivery and end-user-access.

3.1.4 Extraction, Transformation and Loading (ETL)

Data for the data warehouse may come from a variety of data sources (e.g. policy transaction files, claims files, point of sale databases, external data sources). Some data may be incomplete and some may contain errors. Data definitions may be inconsistent among different source data systems. Analytical needs may dictate different levels of granularity and summarization. The ETL process involves accessing the data, staging, cleansing and validating the data, linking data from various systems (e.g. based on account, policy or insured risk), transforming the data to fit various analytical needs (e.g. summarization and deriving measures) and loading the data to the data warehouse.

3.1.5 Tools for end-user access and analysis

The level of access should depend on the expertise and needs of the end-user. Some expert analysts may want to have access to the data in a client/server environment, enabling them to do extensive analyses. Others may prefer to have access over the corporate intranet or the Internet. Vendors often tout the capability of their architecture or products to support a variety of functionality, but much of it may be irrelevant to most users. For example, actuaries rarely if ever need access to data at the policy level with name and address information. The more complex the functionality, the more complex will be the architecture.

3.1.6 MDDB size and access time

For OLAP, an MDDB contains the value of each measure under different combinations of dimensions. The size of an MDDB increases exponentially with the number of dimensions, so it can get very big, very fast. For instance, ten dimensions with eight values each would have 8^{10} (12,058,624) combinations. At some point, an MDDB may become too big to access efficiently, since the software engine must find the answers among all of the cube's data points. Two concepts to consider:

1. Granularity. The level of granularity should be dictated by the needs of analysis. For example, policy or risk-level granularity may not be necessary for most pricing or actuarial applications.

2. Number and content of MDDBs. As with the data warehouse itself, business users prescribe dimensions and hierarchies (the latter must be specified only if the software requires these to be established before the MDDB is created), but the developers determine how best to implement these requirements. The developers must

determine how many MDDBs there should be and what class variables should be handled by each MDDB. Here, redundancy may actually help, since certain factors may be found in more than one MDDB so that those factors can be included in various combinations and hierarchies. The business users must tell the developers the number of values for each class variable, since this affects cube size and efficient cube design. For instance, the variable "state" may have 50 values while "age" (age group) may need to have only four or five bands.

3.2 Case Study: Implementation of the P&C Data Mart

The source data is obtained from a large data warehouse containing transaction-level and account-level records. These records are pre-processed and summarized on the account level and then by the various factors for use as-is for regulatory reports, queries and data mining, and for further processing into MDDB cubes.

The P&C data mart and MDDBs are refreshed monthly, so all the processing described below takes place only once each month (at the beginning of the month).

3.2.1 Factors for analysis: P&C class variables

The business users selected class variables before the start of any programming. They required some changes from the way the factors are represented in the source data. For instance, source data may show the driver ages, but there are too many age values, with too little difference between consecutive values, to be useful. Little is gained in exploratory analyses, by distinguishing between consecutive or near-consecutive values. So ages were grouped into bands (e.g., 16-20, 21-24, 25-29, etc.).

3.2.2 Processing policy, premium and claims data for use in creating MDDBs

The source records included a great number of fields containing policy, customer and vehicle information. Processing the transactional records includes selecting and manipulating fields and using lookup tables to add additional more general information to each record, such as state-specific limitations or the values of certain variables based on rate class codes. End-to-end, the data goes through the following transformations, with the contents of the P&C Data Mart represented by the second, third and fourth boxes in the flow chart:

TRANSACTIONAL RECORDS
↓
ACCOUNT-LEVEL RECORDS
↓
SUMMARIES
↓
MDDB CUBES

These transformations occur one time at the start of the monthly processing cycle, rather than every time information in the data mart is used to generate reports. "Summarization" refers to the aggregation of measures, such as earned premium and incurred claims, by different factors (class variables) and combinations of factors. There may be different levels of aggregation, e.g., year, quarter, and month.

The tables containing account-level records provide a base for regulatory reporting and statistical analysis (regression, data mining). The summaries and MDDBs provide a powerful platform for analyzing the effect of various factors (geography, driver age group, etc.) on premiums and claims. Whether the analyst draws upon the account-level data or summarized data depends on the inquiry. Regression and data mining will probably draw upon account-level data for use in statistical analysis. Summary data will be used for queries that do not require account-level data and for OLAP (after conversion into MDDBs).

3.2.3 Achieving efficiency in monthly processing

The entire process is summarized in Figure 3. In the source data, one record is created each month for each customer to record earned premium, and one or more additional premium records are created whenever one or more policy characteristics (e.g., persons covered, limits, deductible, perils – class variables) change or the policy is renewed. If policy characteristics do not change, then part of the current monthly record is redundant with the previous month's policy characteristic information, which would create processing and storage inefficiencies. The data mart stores this information in a form that is more efficient for both processing and storage by eliminating the aforementioned redundancy. As the monthly source data is processed into quarterly customer records, the repeating policy characteristic information ("policy history") is separated from the premium and claims information. Each policy history record includes the start and end dates to which it applies. The starting date for that record – the date of a policy change or the beginning date of a new policy – is the transaction effective date ("TRANSEFF_DATE"). As an example, if the characteristics of a policy do not change for four months, and is repeated in each of four monthly records, we really need only one policy history record instead of four. This saves storage space, but more importantly it significantly speeds up processing monthly data into quarterly tables. The policy history dataset is in the second row of Figure 3.

Policy information is joined back to the premium and claims data such that the premium and claims amounts are each summarized for the periods during which the policy characteristics have not changed. If there are no such changes during the quarter, then three monthly records become one. The quarterly files contain account-level data not summarized to a higher level (class variables) but with reduced redundancy. They are used for data mining, other statistical analysis, and certain regulatory reports.

3.2.4 Processing monthly transaction files (premium, claims)

Now we begin to examine monthly processing in more detail. Monthly transaction files are identified in the top row of Figure 3. They contain premium and claims data for each type of coverage under each policy. They are created during the processing of source data not shown in Figure 3. (Picture the source data as being in a row above the monthly transaction files.) The monthly transaction files are then summarized into quarterly files and discarded.

The volume of transactions may be very large. For example, for auto insurance, each policy may cover several different types of coverage (BI liability, PD liability, Comprehensive, Collision, Medical, Uninsured Motorist, ADD…). Say the average number of coverage types per policy is 5. If the company has 1 million customers and wants to maintain 6 years of history by month, there would be 360 million such records (1 million customers x 5 coverages x 6 years x 12 mos/yr). This summarization means that the P&C Data Mart uses the quarter as its measure of time, rather than the month. This is usually sufficient for decision support in actuarial applications.

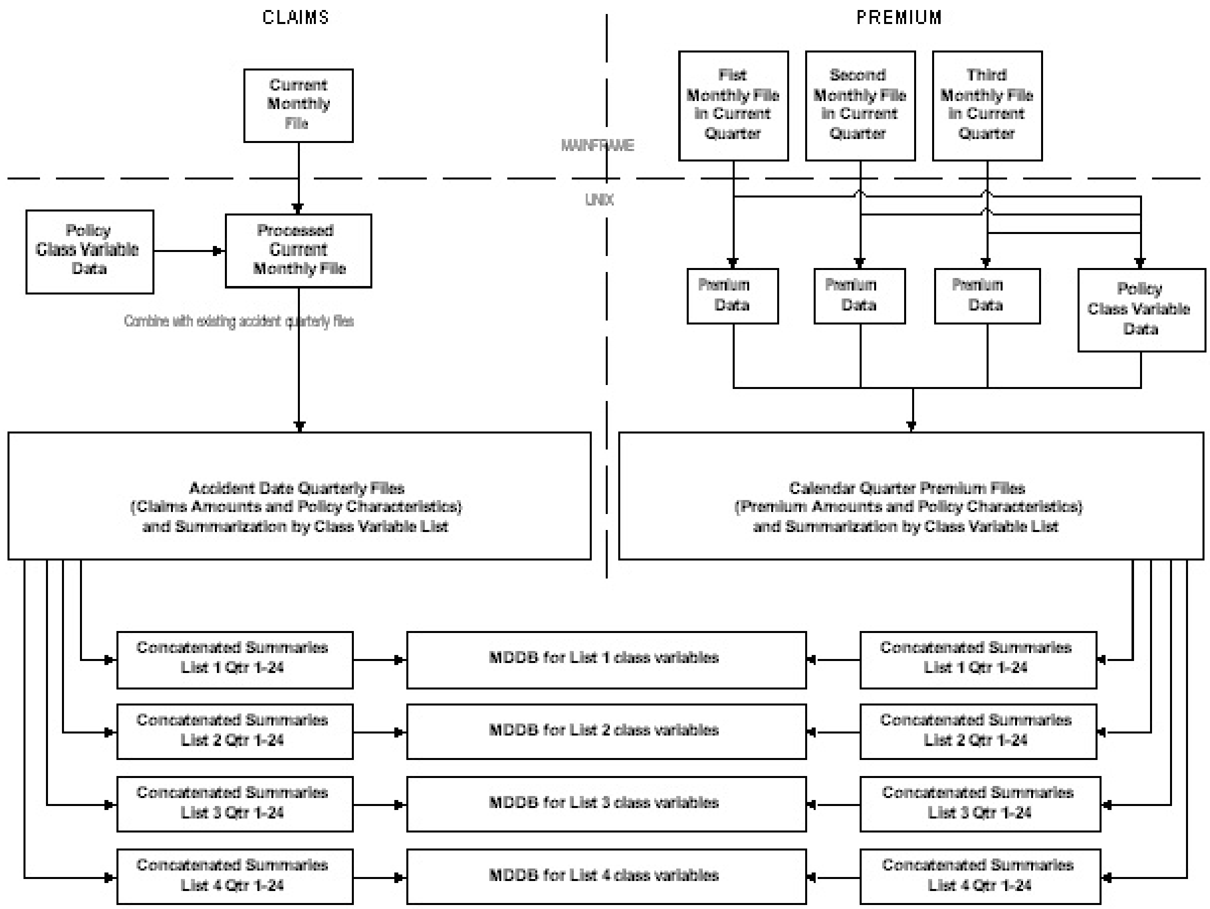


Figure 3: Overview of P&C OLAP Regular Processing

For each claim, the monthly file contains (i) customer ID and peril, (ii) cumulative incurred loss (payments + reserves - subrogation - salvage) starting from the accident date, and expenses, and (iii) deductible/limits as of the accident date. It is critical to understand that the accidents included in the monthly file will have occurred in either the current month or some prior month or even some prior quarter. To append a TRANSEFF_DATE and policy characteristics to each record in the claims file, we use the accident date as a lookup date and match to the policy history file, finding the last TRANSEFF_DATE before the accident date for that policy.

Each of the 24 quarterly claims files contains a record for each accident that occurred in that quarter. Each quarterly claims file is updated every month, with information from those records in the current monthly file regarding accidents in the period covered by that quarterly file. We do not add records to a quarterly file unless the currently monthly file contains the first report of a particular accident in that quarter. Instead, we update records in the quarterly files by combining each record for which the monthly file has new information record with the existing record. For instance, if the June, 2004 monthly file has the record of a payment on a claim for an accident that occurred in the first quarter of 2003, we add the new payment to the claims paid total in the relevant record in the claims file for that quarter (first quarter/2003). Of course, some fields (such as payments) are additive, others (yes/no fields dealing with status) are not, but that is handled by programming.

For each policy, the monthly records for premium and exposure are combined into a new, permanent quarterly file that reflects the total written and earned premium and exposure for the quarter. TRANSEFF_DATE should already be a part of premium records, so we need not append it. For premiums, the accident quarter concept does not apply. Three current monthly files are summarized as the current calendar quarter. The prior 23 permanent quarterly files are untouched. We remove the file for the 24th prior quarter. For processing efficiency, we can store but not process the monthly premium information for the first and second month of each quarter, delaying processing until the third month.

3.2.5 Policy History

This file is represented by the box "Policy Class Variable Data" in the second row of Figure 3. It accumulates the details about the characteristics of each policy as of the transaction effective date, described above. In the case of auto insurance, the key variables for this dataset (what distinguishes one record form another) are POLICY_ID, VEHICLE_NUM and TRANSEFF_DATE. The many fields in each record in this dataset cover policy characteristics (e.g., term, effective date), driver characteristics (e.g., driver education (yes/no), age or age group, points), and vehicle characteristics (e.g., make, model year).

3.2.6 Quarterly files combine claims and premium data (respectively) with policy history

Refer to the third row in Figure 3. Policy history data is included in the records in the quarterly claims and premium files. In auto insurance, the claims and premium data would be matched to policy history by policy identification number (POLICY_ID), vehicle number on the policy (VEHICLE_NUM, say 1, 2, 3, etc., not to be confused with VIN) and the

TRANSEFF_DATE. Here again, claims and premiums are treated very differently: Premium files have many times the number of records as claims files since there may be premiums accrued for each customer each month. Only the current premium quarterly file is affected each month, but all quarterly claims files must be updated (re-created) each month. This is why we keep separate quarterly files for claims and premiums and do not combine the information until after we summarize by policy characteristics (as described below).

Claims. As described in section 3.2.3, all of the quarterly claims files are re-created every month. Fortunately, the volume of claims records is small compared to the volume of premium records, and policy history information must be appended only to the newly arrived claims records for the month. We append policy history to monthly claims data before updating the quarterly claims files.

Premiums. Even at the quarterly level, the number of records in 24 quarterly files may be huge. Fortunately, the quarterly premium files are permanent, since premium is determined for the current quarter only. Only the quarterly premium file for the latest quarter (see 3(b)) must be matched to the policy history. Most customers will have premium records for each of the three months in the quarter, and we summarize such records to aggregate monthly premium. To avoid unnecessary processing, we postpone appending policy characteristics and summarizing monthly records until the third month of the quarter. At the same time, we remove the oldest (24th prior) quarterly file from the data mart.

3.2.7 Summarization by class variable lists and concatenation of quarterly summarized files

MDDBs are manufactured from summaries described in this section. The records in each account level quarterly file are summarized by certain combinations of class variables, referred to here as class variable lists, as prescribed by the users when the data mart is being designed. This moves us from account-level records to the class variable list level.

The particular module of the underlying software system (SAS) used to create MDDBs for the P&C Data Mart required the hierarchies to be selected before creating the MDDBs. (SAS now has additional ways to create MDDBs.) For instance, if analysts will need to drill down from, say, state→age→vehicle use, then we had to summarize the records from each claims quarterly file by that combination, along with any other combination of factors ("class variable lists") that the MDDBs must support. This takes up much of the processing time, but there are efficiency techniques available: (1) Avoiding unnecessary re-summarization of the quarterly premium datasets, and (2) using an intermediate stage of summarization. We will briefly describe these techniques.

Since the quarterly premium files do not change from quarter to quarter (except that we get rid of the oldest quarter and add the file for the quarter just completed), once we summarize one of these files we don't have to do it again unless a new class variable or hierarchy is to be added. So for premiums we only perform this time-consuming group of summarizations once per quarter, and only on one quarterly file (the new one) instead of 24. On the other hand, all 24 quarterly claims files are updated each quarter (see section 3.2.3), so we must re-summarize each of them. Fortunately, the claims quarterly files have only a fraction of the number of records of the premium quarterly files.

Summarization may be made more efficient by doing it in two stages.

Stage 1: Summarize by each of two large collections of the class variables, selected so that class variables that will form hierarchies are included in the same collection. Class variables that are found in many hierarchies may be included in both summarizations.

Stage 2: From the larger summaries, summarize by the actual hierarchies.

If the software engine does not require hierarchies to be established before creating MDDBs, and instead allows users to combine class variables "on the fly" in any hierarchical order, then summarization by specific hierarchies will not be necessary, and larger summaries, with many class variables, could be used instead. Even so, if we do not divide class variables among several MDDBs we are likely to create a single MDDB that is too large to access efficiently.

3.2.8 Concatenate 24 quarterly claims and premium summaries (respectively) for each class variable list

For each list, (i) for claims, concatenate the 24 re-created claims summary files, and (ii) for premium, start with the existing concatenated 24-quarter file, remove records for the oldest quarter, and add records for the current quarter. Result: The number of files = number of class variable lists x 2 (claims and premium). In row 4 of Figure 3, these concatenated summaries are represented by the rectangles on the left and right side. Each such summary file now has information for all 24 quarters.

3.2.9 Merge claims and premium 24-quarter aggregate files

Merge the concatenated claims summaries with the concatenated premium summaries by each class variable list (one file per list). The common characteristics for merge are the class list variables for the corresponding claims and premium concatenated 24-quarter summaries. Now we have brought the 24-quarter claims and premium information together by class variable list. Figure 3 combines this step with the creation of MDDBs (row 4).

3.2.10 Create the MDDB cubes

Each class list becomes a set of class variables for one of the MDDBs. The analysis variables are incurred losses (and the components thereof), premium and allocated expenses. Creating the MDDBs is as simple as using a "Proc MDDB" in SAS, or VB code in the Microsoft universe, or a GUI in either SAS or Microsoft's more recent products.

3.2.11 Create the user interface

As was the case with turning summaries into MDDBs, creating a very user-friendly OLAP interface does not require reinventing the wheel. There are at least three general choices: Middleware tools (e.g., SAS IntrNet), scripts (ASP or JSP pages, using scripting languages such as Perl, VBScript or JavaScript), or canned, customizable graphical interfaces (such as those available from SAS or Microsoft).

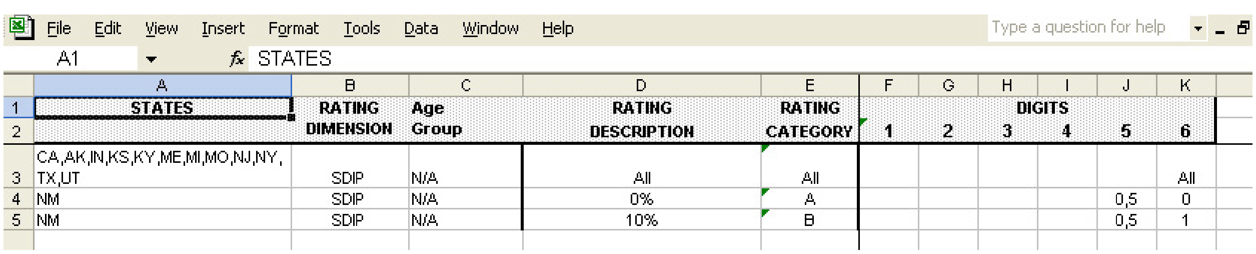
3.3 Reducing the Need for Re-Programming When Business Rules Change

Look for opportunities to eliminate the need for future programming by building flexibility into the system when it is originally being developed. Here's an example of such flexible code included in the code for P&C OLAP System.

The value of ten class variables used in policy rating (pricing) are encoded into a short alphanumeric text string (mostly numeric) called the "rate class code". Here, "class" does not refer to the class variables in the MDDBs. The variables in question are called "rate class code variables," also called "rating dimensions". Examples of rate class code variables are age group, vehicle usage, and safe driver insurance points. Since insurance is a state-regulated business, the business rules governing the rate class codes vary by state, and any given state's premium rating methodology involves hundreds of different rate class codes. Furthermore, the structure of rate class codes varies substantially by state. Therefore, the program code to interpret the rate class code in conjunction with the State and other factors may entail pages and pages of pure complexity. Changing this code requires a programmer and thorough testing after re-coding.

Here is another way. The user enters changes to the business rules for interpreting rate class codes through a familiar front end, such as Microsoft Excel. One of the programs used for monthly processing contains code that interprets the contents of the Excel workbook and, based on what it sees there, can change the interpretation of class codes, even adding a new variable or changing the allowable values of variables.

The programming techniques are described in an article in the proceedings of the 16th Annual Northeast SAS User's Group,[2] and the proceedings of the 29th Annual SAS User Group International.[3] Samples of the user interface are reproduced here showing a few lines of the "main" input worksheet, which prescribes the value of each variable governed by the class code in conjunction with the value of certain other variables including the state (Figure 4), a worksheet describing the characteristics of the class code variables or "rating dimensions" (Figure 5), and a separate worksheet containing the allowable values for each respective rating dimension (Figure 6). The point of identifying such opportunities is to save money in the long run, so the reader who is not a developer is encouraged to review the first of the aforementioned articles, and their developers should be referred to the second.



	A	B	C	D	E	F	G	H	I	J	K
1	STATES	RATING	Age	RATING	RATING			DIGITS			
2		DIMENSION	Group	DESCRIPTION	CATEGORY	1	2	3	4	5	6
3	CA,AK,IN,KS,KY,ME,MI,MO,NJ,NY,TX,UT	SDIP	N/A	All	All						All
4	NM	SDIP	N/A	0%	A					0,5	0
5	NM	SDIP	N/A	10%	B					0,5	1

Figure 4: A Portion of the Rules Definition Sheet in Rate Class Code Interface Workbook

A more general principal statement of the recommendation made in this section 3.3 is to look for ways to separate the description of business rules from the program code, so that users can implement changes in business rules without calling in the programmers. It's not always possible, but it needs to be considered all the time.

4. CONCLUSION

Data warehouses, data marts, OLAP and predictive analytics are essential components of a Business Intelligence system. The data warehouse enables efficient separation of historical data used for analysis from transactional databases. Business needs must drive decisions about the structure and functionality of the data warehouse or data mart. The data warehouse must be well planned for the organization to realize the expected analytical efficiencies.

File Edit View Insert Format Tools Data Window Help

A5 fx Gender_Marital

	A	B	C	D	E	F
1	**RATING DIMENSIONS**	**RAT_DIM**	**4_CHAR**	**LENGTH**		
2	Age Group	AG_GRP	aggp	**1**		
3	Yrs_Experience	YRSEXP	yrs	**2**		
4	Usage	USE	use	**2**		

Figure 5: A Portion of the Rating Dimension Worksheet

OLAP is for exploratory data analysis, not data mining. Deeper analysis requires use

File Edit View Insert Format Tools Data Window Help

B1 fx CODE

	A	B	C	D	E	F	G
1	**SDIP**	**CODE**					
2	0	0					
3	2	2					
4	0-1	D					
5	7-Mar	E					
6	8 and Above	F					
7	41-100%	G					
8	All	H					
9	Collision only	I					
10	Liability only	J					
11	Liabilty+Collision	K					
12	None	L					
13	ZZZ DO NOT USE						

Figure 6: Worksheet Page with Allowable Values and Formats for SDIP

of specialized tools for data mining, using advanced statistical techniques, decision trees and neural networks. Selection of software for data warehousing should take into account the needs and objectives of the overall business intelligence effort.

Even with a well-designed data warehouse and up-to-date software tools for accessing the data, the enterprise must build and nurture its analytical expertise. Actuaries are uniquely positioned to take a leadership role in maximizing the benefits of Business Intelligence tools. Furthermore, however much we may enhance our capabilities with technology, we must never lose sight of the importance of ingenuity, creativity and a solid sense of the business in analysis and decision making.

Abbreviations and notations
I/O, Input/Output
MDDB, Multi-Dimensional Database
OLAP, On-Line Analytical Processing
TRANSEFF_DATE, Transaction Effective Date (date of change in one or more policy characteristics, or the starting date of a policy.

Biographies of Authors

George Bukhbinder is the President, Palisades Research, Inc. Mr. Bukhbinder has over 20 years of experience database and data warehouse design and development, Internet-enabled information delivery for decision support, statistical modeling and data mining. He has worked extensively on the development and implementation of information systems in the insurance, financial, telecommunications and pharmaceutical industries. Mr. Bukhbinder has a Ph.D. in Statistics. He is a regular speaker at regional and national conferences.

Michael Krumenaker is Sr. Project Manager at Palisades Research, Inc. He has over six years as a full-time programmer in SAS, VBA and C++. Before that he spent seventeen years in corporate finance, including application development in spreadsheets and databases. He has degrees in business (MBA - Columbia University), law (JD - Vanderbilt, LLM - NYU), and chemistry (BS - Brooklyn), and has completed all courses for MS in Computer Science at the New Jersey Institute of Technology.

Abraham Phillips Abraham Philips is an Insurance Industry Consultant. He has an M.S. in Information Systems from Stevens Institute of Technology and a Ph. D. in Statistics from University of Windsor, Canada. He has over 25 years of experience in the Insurance industry conducting and managing data analysis, statistical research, personal lines pricing and actuarial information support including the development of data warehouses, data marts and analytical tools.

5. REFERENCES

---

[1] A database design that is based on a central detail fact table linked to surrounding dimension tables. Star schemas allow access to data using business terms and perspectives. University of Illinois on-line decision support glossary, http://www.ds.uillinois.edu/glossary.asp.

[2] M. Krumenaker and J. Bhattacharya, User Implementation and Revision of Business Rules Without Hard Coding: Macro-Generated SAS Code, Proceedings of the 16th Annual Northeast SAS User Group Conference, paper AD003 (2003).

[3] G. Bukhbinder and M. Krumenaker, Developing Data Marts an Web-Enabled OLAP for Actuarial and Underwriting Analysis, Proceedings of the 29th Annual SAS Users Group International, paper 111-29 (2004).

Dancing With Dirty Data

Methods for Exploring and Cleaning Data

Louise A. Francis, FCAS, MAAA

Abstract

Motivation. Much of the data that actuaries work with is dirty. That is, the data contain errors, miscodings, missing values and other flaws that affect the validity of analyses performed with such data.

Methods. This paper will give an overview of methods that can be used to detect errors and remediate data problems. The methods will include outlier detection procedures from the exploratory data analysis and data mining literature as well as methods from research on coping with missing values. The paper will also address the need for accurate and comprehensive metadata.

Conclusions. A number of graphical tools such as histograms and box and whisker plots are useful in highlighting unusual values in data. A new tool based on data spheres appears to have the potential to screen multiple variables simultaneously for outliers. For remediating missing data problems, imputation is a straightforward and frequently used approach.

Availability. The R statistical language can be used to perform the exploratory and cleaning methods described in this paper. It can be downloaded for free at http://cran.r-project.org/.

Keywords. data quality, data mining, ratemaking, exploratory data analysis.

1. INTRODUCTION

The frequency of poor data quality is one of the most vexing problems for actuaries. Countless hours are expended detecting data problems, remediating the problems and revising analyses after data problems have been revealed. Data quality problems are not unique to the insurance industry. Olson describes data quality problems as nearly universal (Olson, 2003). In his words "In just about any organization, the state of information quality is at the same low level"[1]. Olson cites two explanations for this unfortunate situation: 1) the pervasiveness of rapid system implementation and change and 2) methods for assuring data quality have not developed nearly as rapidly as the ability to collect, store and process data. Olson estimates that the cost of data quality problems is 15% - 20% of operating profits[2].

Insurance companies collect vast amounts of data and use this data to make key decisions, such as the price to charge for an insurance policy and the amount of liability for claim obligations that will appear on the company's financial statements. These data driven decisions are key to the profitability of insurance companies. Insurance companies often

---

[1] Olson, p10.
[2] Olson, p9.

aggregate much of the useful detail out of their data. Pricing and reserving functions are performed on large groupings of businesses. Thus, there are missed opportunities to improve the company's profitability through better use of its data. As a consequence, such efforts to assure data reliability as do exist are focused primarily on assuring the accuracy of large aggregates of financial amounts. For example, it is considered important that the incurred and paid losses that are allocated to a given accident year at a given valuation date for a given line of insurance are correctly stated (even though all too often they are not). Almost no attention is focused on assuring that the values recorded for infrequently referenced fields, such as injury type or return to work date, are complete or accurate. As a result, an insurance company may not be able to monitor the effectiveness of new initiatives, such as programs which aim to return injured workers to work sooner than in the past. These data quality issues also hamper effort to build analytical models focused on finding complex patterns in data, such a fraud analysis requiring injury and treatment information.

In recent years, insurance companies have started to explore the use of advanced analytical techniques in order to more accurately price and reserve the insurance exposures, as well as predict fraud, model catastrophes and other unique exposures, market policies and support other management decisions. These analyses make heavy demands on data and typically involve large databases – often millions of records and hundreds of variables. The data quality problems are magnified for these large scale analytical projects. This is because the projects utilize data not frequently used for other business functions and therefore data quality issues become a challenge to the modeling effort. Analysts devote significant resources to finding, fixing or otherwise remediating data problems. A rule of thumb is that more than 80% of the time devoted to analytical projects is expended on processing and cleaning up messy data (Dasu and Johnson, 2003).

In this paper a number of methods will be presented which can be used to detect and remediate data quality problems. The focus will be on two areas: 1) detecting data errors and 2) finding and adjusting for missing data. The methods presented in the paper are focused on projects using large databases, but they may also be applied to databases of more modest size.

1.1 Research Context

The actuarial literature on data quality is relatively sparse. The American Academy of Actuaries (AAA) standard of practice #23 on data quality provides a number of important guidelines for assuring the validity of data when performing an actuarial analysis. The standard provides guidelines for reviewing data for completeness, accuracy and relevance to the analysis. The Casualty Actuarial Society (CAS) committee on Management Data and Information and the Insurance Data Management Association (IDMA) also produced a white paper on data quality (CAS committee on Management Data and Information, 1977). The white paper states that evaluating the quality of data consists of examining the data for:

- Validity,
- Accuracy,
- Reasonableness,
- Completeness.

These same concerns apply to data supplied to an analyst performing a large analytical project. A typical actuarial review of data consists of balancing totals from the data to published financial reports and inspecting the data for obviously erroneous values, such as negative amounts for financial variables like paid losses. The data quality white paper describes a number of more extensive activities that could be performed to assure the overall integrity of the data systems serving all the different business users within an insurance company. These include data edits to detect impermissible values in the data and periodic data audits to measure the extent of data quality problems.

This paper is focused on data quality issues arising when data is supplied by an external (or internal) source not under the control of the analyst that must be screened for data quality problems prior to use in a project.

A somewhat extensive literature that is relevant to data quality exists in statistical journals and publications. This includes the tools of exploratory data analysis, pioneered by Tukey (Hartwig and Dearing, 1979 discuss Tukey's contribution), and graphical analysis of data, popularized by Chambers and Cleveland (Chambers et al., 1983, Cleveland, 1993). Exploratory data analysis techniques are particularly useful for detecting outliers. While outliers, or extreme values, may represent legitimate data, they are often the result of data glitches and coding errors.

Another aggravating data quality issue is that of incomplete or missing data. In recent years the literature on methods for remediating missing data has been growing. Rempala and Derrig (Rempala and Derrig, 2003) presented the expectation maximization procedure for estimating missing values. Francis (Francis, 2003) described how the MARS data mining procedure creates surrogate variables to use when values are missing. This paper will not cover the EM or MARS approaches, but will review several of the most common methods for "plugging in" values where data is missing. Some of these methods, such as replacing a missing value with the mean of that variable, have been used for decades while others such as data imputation have been developed more recently. In this paper, procedures for detecting and remediating missing data problems will be presented.

1.2 Objective

Data quality is a ubiquitous and daunting problem for analysts of insurance data. A goal of this paper is to raise awareness of the data quality problem in the insurance industry. Because the users of insurance data will frequently be required to do the best they can with data that has quality issues, this paper present some methods for screening data and detecting data problems. Only a few key exploratory and data cleaning methods will be presented in this paper, but the reader is referred to literature in the references section of this paper for further information. Dasu and Johnson (Dasu and Johnson, 2003), in particular, provide a more thorough introduction to procedures that include those in this paper and cover a large number of other approaches, which can be easily implemented.

Many of the exploratory methods presented in this paper are intended to detect outliers, or erroneous values. Missing data is also an important data quality issue; therefore this paper presents methods for detecting and remediating missing data.

1.3 Outline

The remainder of the paper is structured as follows.

- Section 2 is the background and methods section.
  - Section 2.1 will introduce the data set used to illustrate the methods in this paper.

  - Section 2.2 will discuss methods for detecting unusual values in quantitative data. This section will present some well-known visual and numerical summaries of data, which can be used to detect unusual values. It will also introduce the concept of data spheres.
  - Section 2.3 will present methods for detecting unusual values in categorical data. This section will introduce the concept of data cubes. It will illustrate the exploration of categorical data with tabular summaries of the data.
  - Section 2.4 will present methods for finding and remediating incomplete data.
  - Section 2.5 will discuss inappropriate use of insurance data that can arise when censorship, or the presence of incomplete data, is not considered.
  - Section 2.6 discusses metadata.
- A summary of the paper's results and conclusions will be presented in Section 3.

2. BACKGROUND AND METHODS

Inaccurate and incomplete data are universal problems for data analysts. Methods for detecting inaccurate data have existed for many years but are not widely used in the actuarial profession. Methods for addressing incomplete data have also been incorporated into statistical software for many years. However, recent advances have significantly improved the arsenal of tools available for addressing this issue. This paper will illustrate the use of exploratory techniques for detecting data problems and missing values techniques for remediating incomplete data.

A sample database has been created to illustrate the data exploration and data cleaning procedures presented in this paper. The example is based on a sample of actual data used for a large data analysis project, but original values in the data have been modified. The size of the sample data, approximately 35,000 records, is considerably smaller than that used in many large-scale analyses, but its size allows the illustration of many useful techniques for exploring and cleaning data.

2.1 An example using personal auto data

In order to provide an example of data exploration and data cleaning approaches, a 35,284 record database of personal automobile insurance policies was created. The data is representative of data utilized for an actual analysis; however the example data is somewhat smaller in size than data used in an actual large-scale analytical project. The data are intended to be representative of policy and claims data encountered in the personal automobile line of business. Each record represents data for a policyholder. The data elements presented below could be used for underwriting, ratemaking and other insurance applications. The fields in the data are:

Date of birth

License date

Age

Number of vehicles

Number of drivers

Marital status

Territory

Vehicle symbol

Model Year (of the vehicle)

Class code

Business Type (New, Renewal, Targeted or preferred)

Policy type (Liability, Liability and Physical damage, Physical damage)

Policy inception date

Number of claims

Incurred losses

Paid losses

Paid allocated loss adjustment expenses

Ultimate claims

Ultimate losses and expenses

Subrogation

Earned premium

Written premium

Earned exposures

Written exposures

Zip Code

Two major kinds of variables occur in the automobile insurance data 1) categorical (or alphanumeric) variables and 2) quantitative (or numeric variables). Each value on a categorical variable conveys qualitative information that is useful in describing characteristics of the policyholder or classifying the policyholder into one of a number of categories. Examples are gender and the territory where the policyholder's car is garaged. However, the values on a categorical variable, such as "female" or "male" do not have any numeric or ordinal information. On the other hand, quantitative variables such as driver age or paid losses contain quantitative content. An age of 50 years is greater than age of 20 years, and it is greater by 30 years. Losses of $10,000.00 are greater than losses of $100.00 and exceed them by 10,000%. The numeric variables not only convey ordinal information, but measure relative relationships (it matters which one is higher and by how much). Different techniques are utilized to explore and clean the different kinds of variables. Some of the most commonly used of the techniques are described below.

2.2 Numeric variables

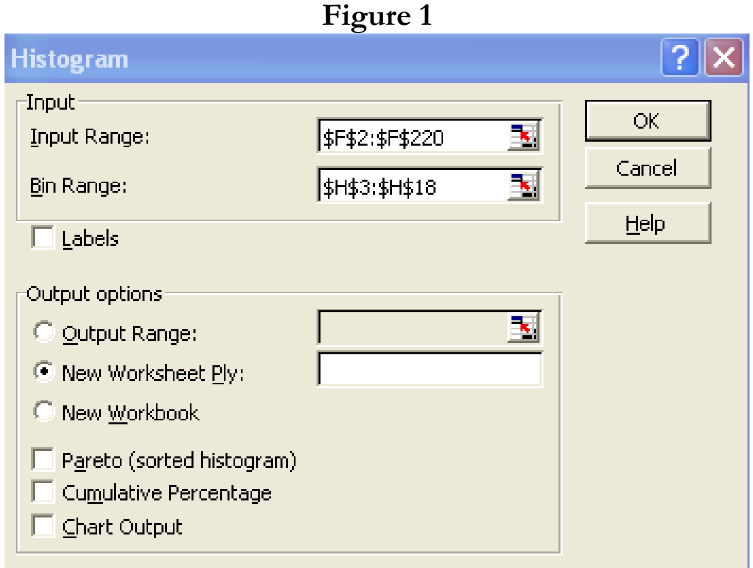
Common errors with numeric data include negative values in financial fields that can have only positive values and values that exceed the possible range for that variable, such as a driver age of 10 in a state where the minimum age for driving is 16. Such errors often appear as outliers, i.e., as extremely small or large values that are outside the range of most of the data. A number of graphical displays assist in the detection of outliers. Once an outlier is determined to exist, it can be investigated and its validity determined. In insurance data, legitimate extreme values are a fairly common occurrence. For instance, because insurance loss distributions are heavy-tailed, extreme values, of more than 3 standard deviations from the mean of the distribution, occur far more frequently than would be expected if data were normally distributed.

Two very useful graphical tools are discussed below: histograms and box and whisker plots.

2.2.1 Histograms

According to Chambers et al., "There is no single statistical tool that is as powerful as a well chosen graph"[3]. Often graphical summaries of data are very revealing and helpful in detecting outliers. One of the most commonly used and understood graphical summaries of the values of numeric variables is the histogram. The capability of producing histograms is widely available. For instance, using Microsoft Excel's data analysis toolpak, a histogram can be easily created. The user specifies a bin range and the column of data for which a distribution is being created (see Figure 1). For instance, Table 1 presents the bin ranges, which might be specified for the driver age variable. The bin ranges specify the intervals that the data is grouped into. Since the first interval in Table 1 is 20, the total count of drivers with ages less than or equal to 20 will be summarized in the first bin. The second bin interval is 25, so the number of drivers with ages greater than 20 and less than or equal to 25 will appear in that bin. Once the count of records in each bin is summarized, a graph of the distribution of records in each interval can be created. The y-axis of the graph generally displays either the total count of records in the interval or the percent of total records in the interval. It is common to select evenly spaced intervals, but there are occasions where varying bin widths are preferable.

Figure 1

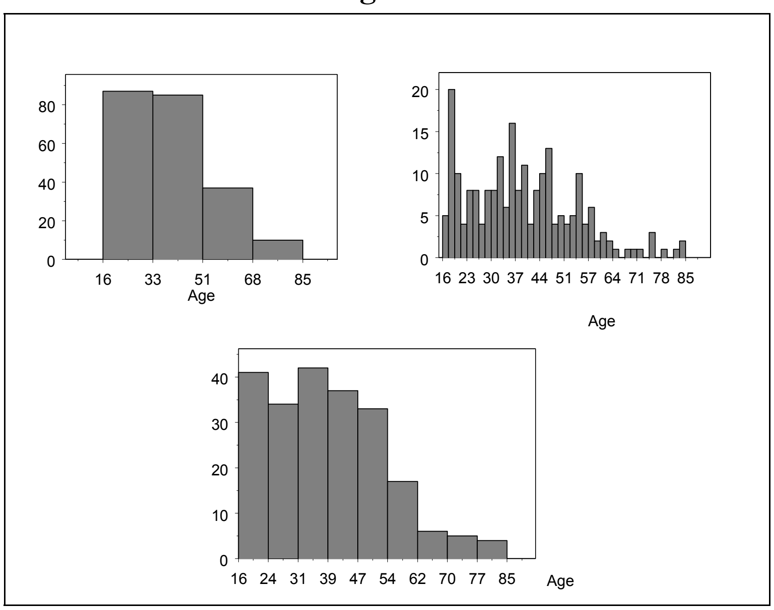


[3] Chambers et, al., p1.

Table 1

Bin	*Frequency*
20	2853
25	3709
30	4372
35	4366
40	4097
45	3588
50	2707
55	1831
60	1140
65	615
70	397
75	271
80	148
85	83
90	32
95	12
More	5

Figure 2



Three histograms of age for a sample of 220 records. In Figure 2, the top left illustration has four bins and the top right graph has 40 bins. The bottom figure has 9 bins as determined by equation 2.1.

A relatively small sample from the automobile insurance database was used to produce the histograms in Figure 2, in order to illustrate issues relating to how the records underlying the graph are grouped. When a small number of bins (wide bins) is selected, a much cruder image is created. However too many bins may result in a noisy image, and makes the overall shape of the distribution difficult to determine. A rule for selecting the width of the histogram bins (also known as window width) is (Venibles and Ripley, 1999):

$$h = \frac{3.5\sigma}{\sqrt[3]{N}^{\frac{1}{3}}} \qquad (2.1)$$

h is window width
σ is the standard deviation of the variable
N is the number of records

This window width rule was derived under the assumption that the data has a normal distribution. For the data in Figure 2 (a sample of 220 records), with a standard deviation of 15, the rule yields the following window width:

$$h = \frac{3.5*15}{\sqrt[3]{220}^{\frac{1}{3}}} = 8.7 \qquad (2.2)$$

By dividing the range of values (the maximum value minus the minimum value) by the window width h, the number of bins can be determined. The range of values in the data is 84 (100 – 16). Dividing this by the window width of 8.7 yields between 9 and 10 intervals.

The formula above provides a rule for determining the window width for equi-spaced histograms. An alternative to an equi-spaced histogram is an equi-depth histogram. (Dasu and Johnson, 2003). In an equi-depth histogram the same percentage of records are used for each bin, therefore each bin contains the same mass.[4]

[4] In using equi-depth bins, the analyst might wish to divide by the bin width, creating a meaningful measure of density. This would avoid having all the bars the same size.

Figure 3

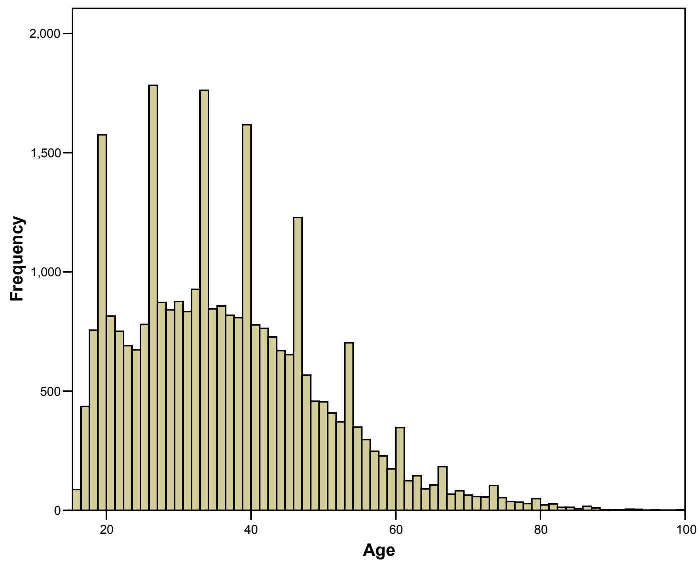


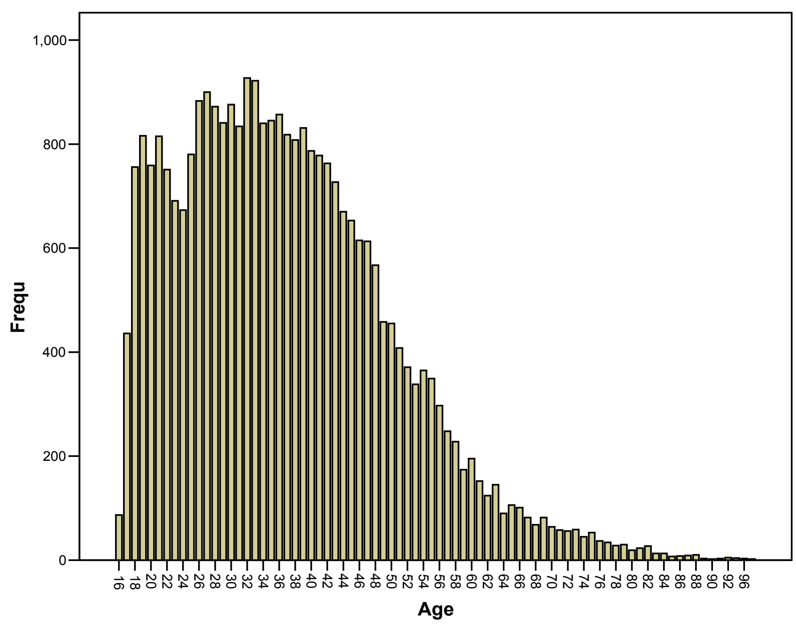
Figure 3 presents a histogram of the age variable, for the full data set of about 35,000 records. It should be noted that many of the widely available statistical packages default to a rule such as equation (2.1) for determining the number of bins to use for grouping records used in a histogram[5]. Figure 3 shows that there are a few very old drivers in the data. The analyst might wish to investigate the validity of these extreme values.

The figure also indicates that there are periodic jumps in the frequencies of age. Some survey research (Carter and Bradley., Heitjan) indicates that ages are sometimes under or over-reported and that "rounding" of ages may occur. That is, there may be some rounding at certain ages, such as ages ending in 0 or 5. The age data was examined in more detail for systematic patterns indicating underreporting or over-reporting at some ages. Figure 4 presents the graphical results of examining the age data in greater detail. This graph, which shows the frequency of records at every age reported in the data, displays no large jumps. A more careful review of the binning procedure resulting from application of formula (2.1) indicates they applying the rule to ages reported in years results in periodic grouping of the frequencies for two years together, roughly doubling the counts compared to the

[5] Note that most statistical software automatically selects the scale (minimum and maximum values for each axis) as well as the number of bins using default rules. The users are allowed to choose other options if they do not like the default rules.

surrounding ages. Thus, the analyst needs to exercise care when applying of any rule for binning data, as features specific to that variable can produce unexpected results.

Figure 4



The next example illustrates an instance where the histogram helps to detect an obvious data glitch. A histogram of the variable license year is presented in Figure 5. The graph takes an unusual shape: most of the observations are clustered in the right hand of the graph, but a very small percentage of the mass lies in the extreme left. It can quickly be surmised from the graph that at least one record contains erroneous values on this variable, i.e., a license date that is prior to the year 600. To find the outlier observation(s), the data was sorted by ascending order on the license year variable. Table 2 presents the 18 lowest observations on this variable. All have license years prior to 1900, clearly impossible values.

Figure 5

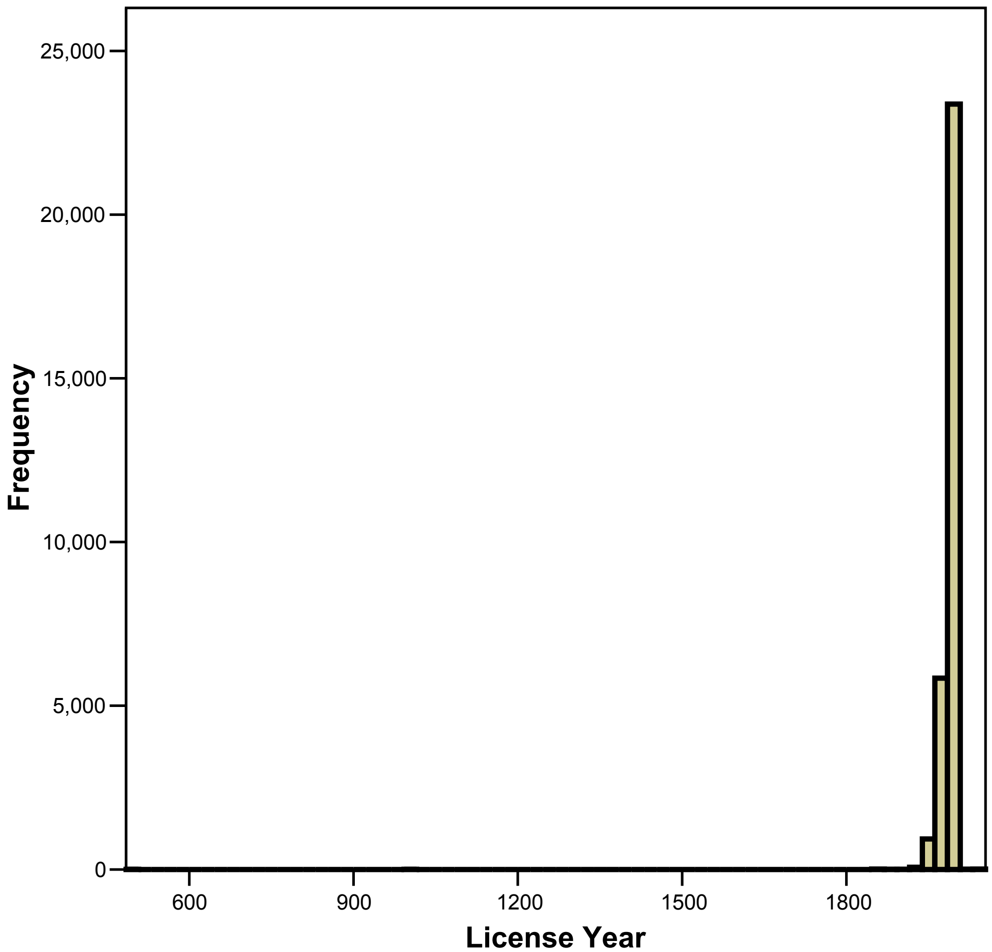

25,000
20,000
15,000
10,000
5,000
0
Frequency
600
900
1200
1500
1800
License Year

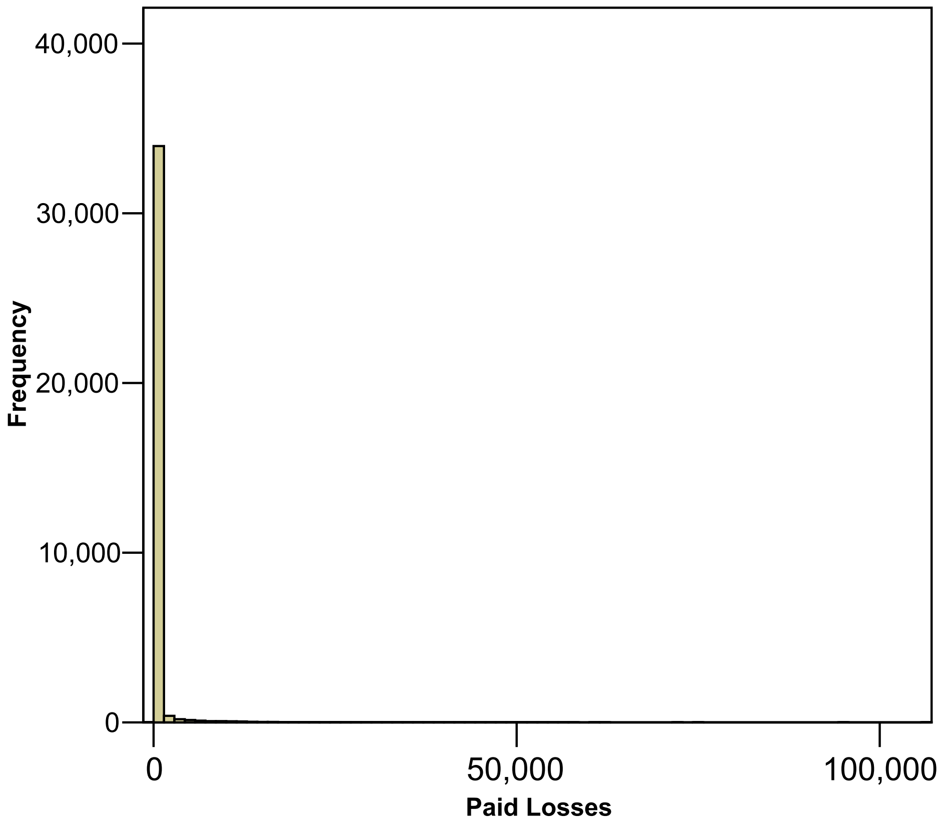
Table 2

Policy ID	Licensed Date	Licensed Year	Date of Birth	Birth Year	Age
28319	7/1/0490	490	7/4/1972	1972	30
08861	2/1/1000	1000	12/31/1983	1983	20
00043	1/1/1857	1857	7/19/1966	1966	36
01203	1/1/1857	1857	8/21/1965	1965	38
02003	1/1/1857	1857	10/14/1975	1975	28
03132	1/1/1857	1857	6/6/1947	1947	56
04114	1/1/1857	1857	5/21/1961	1961	42
04839	1/1/1857	1857	8/28/1970	1970	33
05338	1/1/1857	1857	10/3/1978	1978	25
05339	1/1/1857	1857	10/3/1978	1978	25
05424	1/1/1857	1857	2/23/1949	1949	54
05946	1/1/1857	1857	6/22/1976	1976	27
06028	1/1/1857	1857	9/13/1980	1980	23
06175	1/1/1857	1857	2/16/1965	1965	38
06386	1/1/1857	1857	5/27/1980	1980	23
34079	1/1/1857	1857	8/21/1965	1965	39
34930	1/1/1857	1857	10/2/1985	1985	19
04342	6/19/1890	1890	6/19/1963	1963	40

The license date value for all records with a value below 1900.

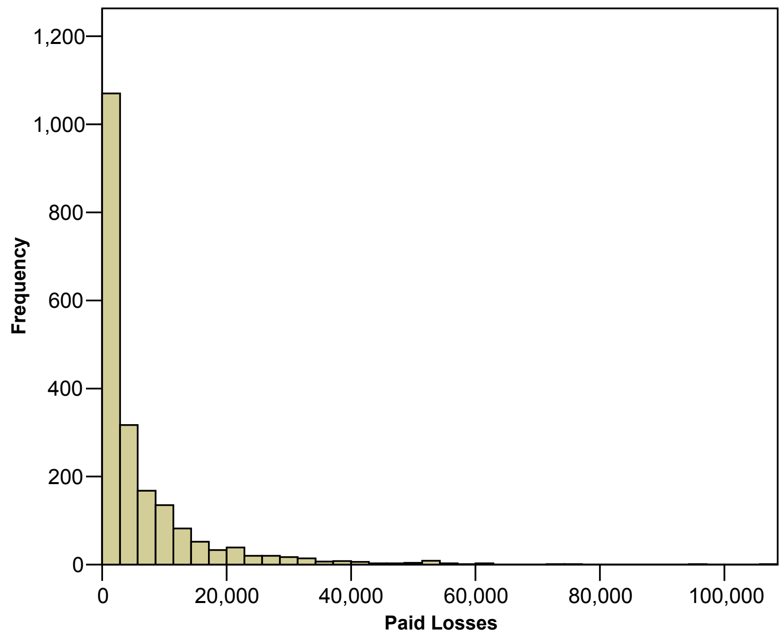
Figure 6 illustrates another issue that arises when visually screening data with histograms. This graph displays a distribution for paid losses (per policyholder, not claimant) where the number of bins has been determined according to equation 2.1. For this graph, the overwhelming majority of the records are displayed on the left side of the graph at the origin, with almost no perceptible mass at other values. This occurs because approximately 90% of the records are those of policyholders who have not reported a claim; therefore there is a large mass point for the histogram at zero. This histogram is relatively uninformative with respect to drawing useful conclusions about the important characteristics of the paid loss distributions. One approach to dealing with data that has a mass point at zero is to filter the paid loss data and remove from the graph those observations with a zero value. When filtering data, records we are not interested in are removed from the statistics and charts being produced. However, the records remain in the data for use on other procedures and charts. Many analytic tools, including Microsoft Excel, provide the user with the capability of filtering data.

Figure 6



Histogram of paid losses including all records

Figure 7



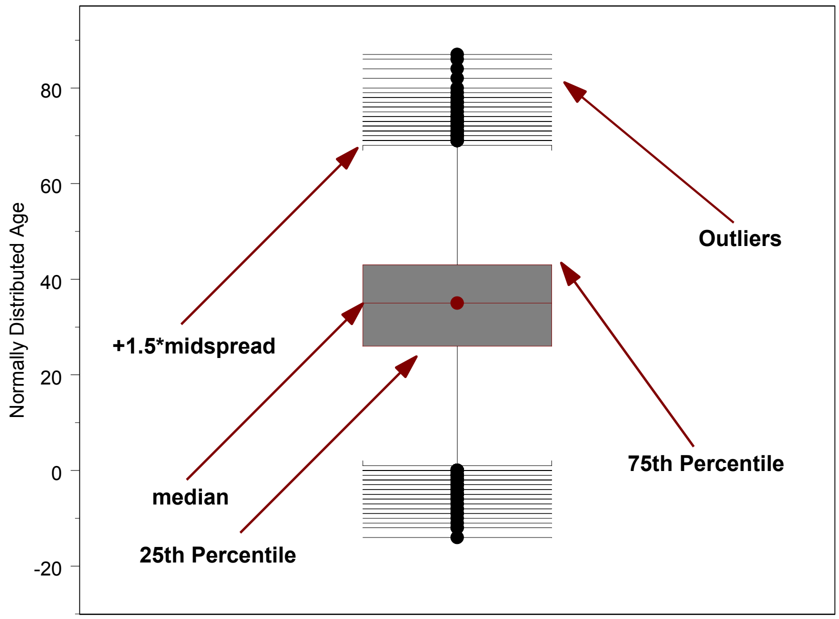
Histogram of paid losses for filtered data.

The features of the paid loss data displayed in Figure 7, based on filtered data, are more informative than those of Figure 6. At this point nothing seems amiss with this data. There appear to be many records with relatively modest paid loss amounts and a few records with large amounts, but nothing that is unexpected or unusual for paid loss amounts. In the next section a procedure is presented which highlights key features of a distribution that may not be obvious from the histogram.

2.2.2 Box and whisker plots

One of the most useful graphical displays for exploring and cleaning data is the box and whisker plot first introduced by Tukey. The box and whisker plot provides a one-dimensional summary of key features of numeric data. The basic components of the box and whisker plot are illustrated in Figure 8. The key components of the plot are 1) a box, 2) two whiskers extending from the box and 3) outliers.

Figure 8

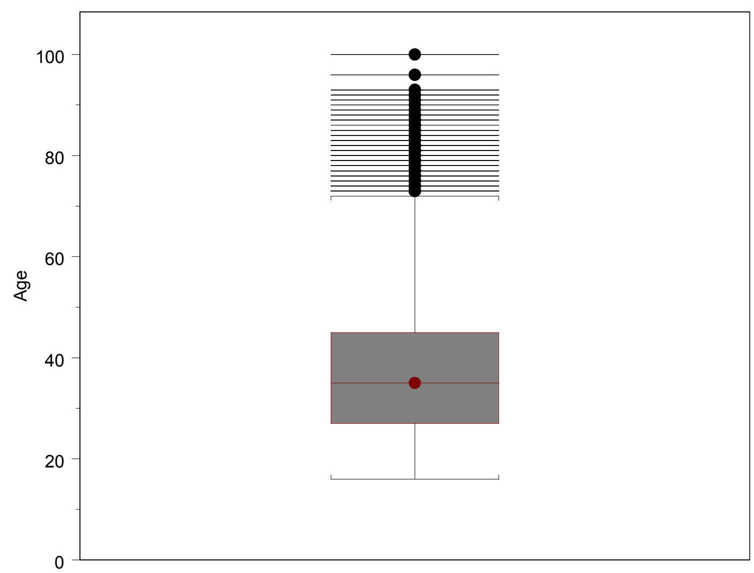


Key features of the box and whisker plot are the median, the edges at the 25th and 75th percentiles, the whiskers and the outliers

The (edges) top and bottom of the box are defined by the 75th and 25th percentiles of the distribution plotted. A line through the middle of the box denotes the 50th percentile or median value. (The width of the box carries no meaning). A line extends both from the top and bottom of the box. These lines are referred to as the whiskers. For this graph, the lines denote the points 1.5 midspreads above and below the box edges (the midspread is the difference between the 75th and 25th percentile). Different rules can be utilized to determine the length of the whiskers. Another rule commonly used is for the whiskers to have a length 1.5 or 2 times the standard deviation of the distribution. In Figure 8, points beyond the 1.5 times the midspread boundary are individually displayed (the circles with lines through them). These points may be considered outliers. The points denoted as outliers depict records that the analyst might want to investigate.

Figure 9 displays the box and whisker plots for the age field in the auto data. This data is not normally distributed. Because the data is right skewed, the whisker for the upper portion of the distribution is larger than the whisker for the lower portion of the distribution. Moreover, only the right tail displays extreme values. This graph, like the histogram, indicates there are some records with very high values that an analyst might want to investigate.

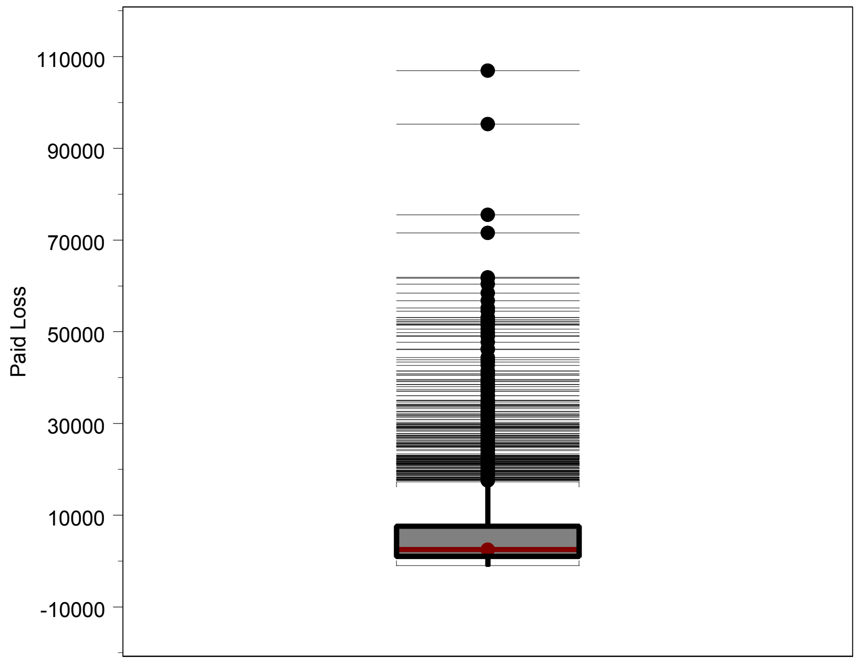
Figure 9



Box and whisker plot of age from auto data

Figure 10 displays a box and whisker plot for filtered paid loss data (that is, the paid losses were filtered to remove zero values). Because the distribution of paid losses is very heavy tailed, the top whisker is much longer than the bottom whisker. In addition, the box enclosing the interquartile range is extremely small and important statistics such as the median of the data cannot be read from the graph. While the many circles at the top of the graph indicate a relatively large number of extreme values, such values are normal for insurance financial variables. A more useful plot with more ability to identify real outliers could be constructed on rescaled or transformed data. In order to make the graph interpretable, a display on a log scale using a base of 10 is reasonable.

Figure 10



Box and whisker plot of paid losses. Data are on untransformed scale.

The box and whisker plot displayed in Figure 11 provides a much more interpretable summarization of the paid loss distribution than Figure 10. If an error is introduced by introducing a value well outside the range of the data (in this case the paid losses on one record was recoded to $10 million), the box and whisker plot can be used to detect the outlier. This is shown in Figure 12, where a point is plotted at the top of the graph, which is orders of magnitude higher than all the remaining data.

Figure 11

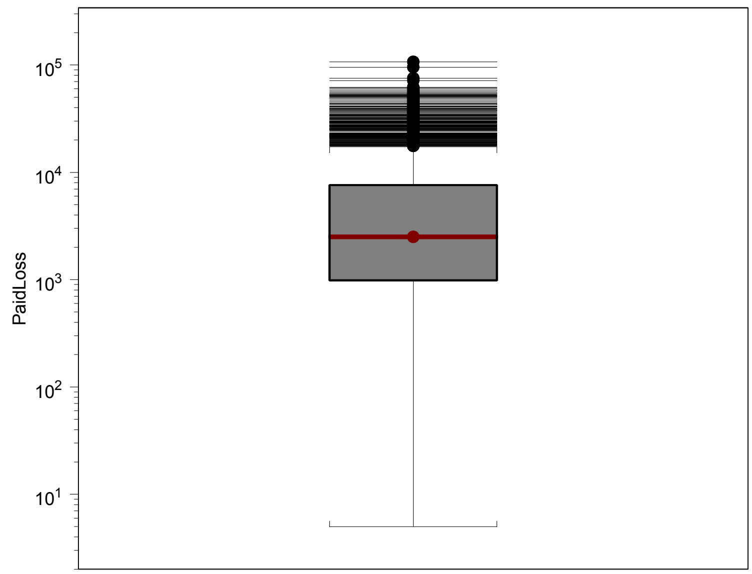
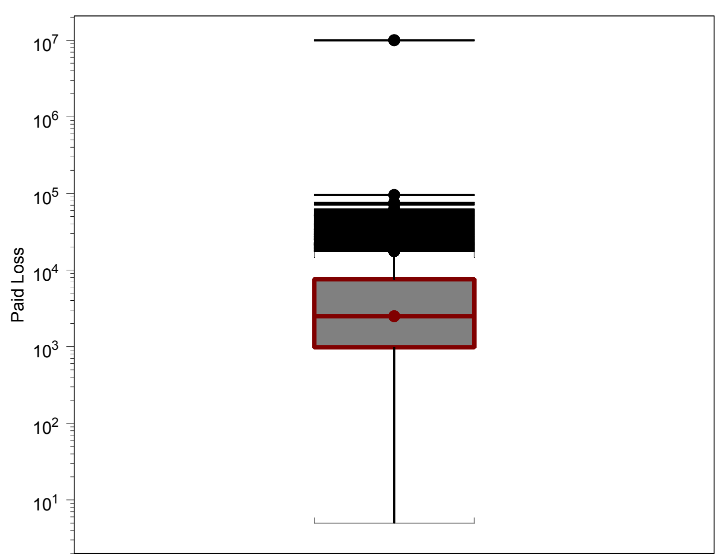


Figure 12



Box and whisker plot without (Figure11) and with (Figure 12) outlier on paid loss data.

2.2.3 Descriptive statistics

A quick way to screen numeric data for invalid values is to produce summary tables of descriptive statistics. Such tables can usually be quickly prepared using commonly available statistical packages. Descriptive statistics output displays the most important statistics characterizing a distribution. Some of the most common statistics displayed are the mean, median, minimum and maximum. The analyst can quickly review the descriptive statistics tables for an indication that the data contain inappropriate values.

Table 3 displays descriptive statistics for the license year variable. The statistics for the minimum and maximum values both indicate problematic values for this variable. Table 4 displays descriptive statistics for age and indicates a policyholder of age 100 years, an extremely high value for this variable. Table 5 presents descriptive statistics for the paid loss variable. The minimum for this variable indicates a suspicious (negative) value. These are three examples of how simple summaries of numeric variables may give an indication of unusual values.

Table 3

	N	Minimum	Maximum	Mean	Std. Deviation
License Year	***30,250***	***490***	***2,049***	***1,990***	***16.3***
Valid N	***30,250***				

Descriptive statistics for license year

Table 4

	N	Minimum	Maximum	Mean	Std. Deviation
Age	***30,242***	***16***	***100***	***36.9***	***13.2***
Valid N	***30,242***				

Descriptive statistics for age

Table 5

	N	Minimum	Maximum	Mean	Std. Deviation
Paid Losses	***35,284***	***-1,000.00***	***106,940.00***	***364.57***	***2769.8***
Valid N	***35,284***				

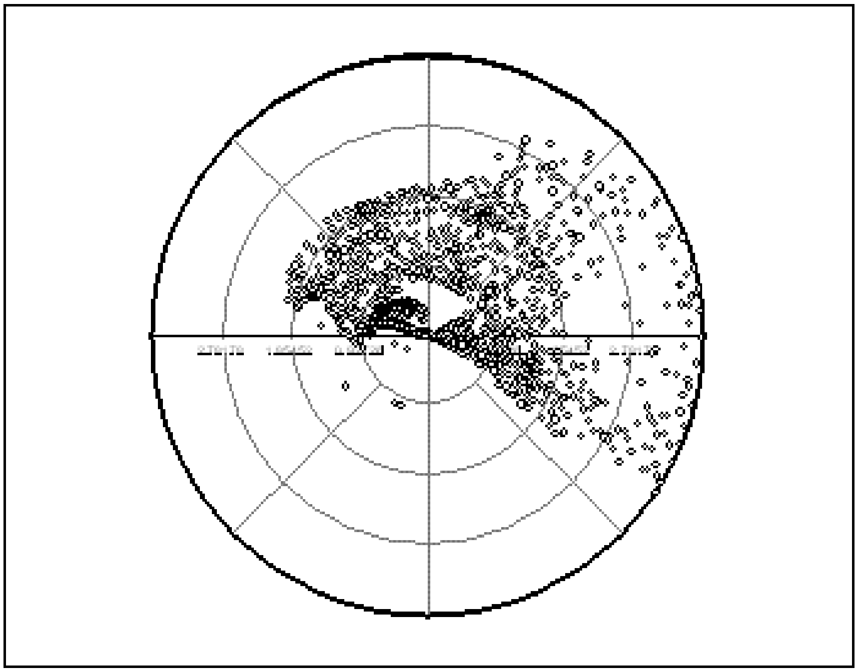
Descriptive statistics for paid losses

2.2.4 Data Spheres

The data exploration methods described above are based upon screening variables one at a time. Dasu and Johnson (Dasu and Johnson, 2003) recently introduced the concepts of data spheres for simultaneously screening a number of variables for outliers. Their logic is that records with typical values for data are near the "center" of the data and records containing outliers are a large distance from the "center" of the data.

To illustrate the concept of data spheres, a plot was created using the latitude and longitude for the zip code associated with each record. This information was obtained by incorporating into the original data, geographical data obtained from a third party vendor. The latitude and longitude data were standardized so that the mean of each variable is zero and the standard deviation is one. Figure 13 displays a circular plot of the latitude and longitude data. This plot indicates that most of the records lie within the 2nd innermost circle of the data, but a few points lie along the perimeter. Those points along the perimeter represent geographic outliers. In fact, the tabulation of records in Table 6 indicates that most policyholders are located in one state, but a small percent are in other states.

Figure 13



Circular plot of latitude and longitude

Table 6

State		Frequency	Percent	Valid Percent	Cumulative Percent
		26	*.1*	*.1*	*.1*
	CA	*1*	*.0*	*.0*	*.1*
	FL	*2*	*.0*	*.0*	*.1*
	MA	*2*	*.0*	*.0*	*.1*
	NC	*1*	*.0*	*.0*	*.1*
	NJ	*9*	*.0*	*.0*	*.1*
	NY	*3*	*.0*	*.0*	*.1*
	PA	*35,240*	*99.9*	*99.9*	*100.0*
	Total	*35,284*	*100.0*	*100.0*	

Dasu and Johnson (Dasu and Johnson, 2003) introduce the Mahalanobis depth as a way to measure how far a given record is from the center of the data. The statistic is:

$$\mathbf{MD} = (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \tag{2.3}$$

where $\mathbf{x}$ is a vector of variables, $\boldsymbol{\mu}$ is a vector of means of the variables, $\boldsymbol{\Sigma}$ is the variance-covariance matrix of $\mathbf{x}$

This formula indicates that the Mahalanobis depth measures the squared deviation of each variable on each record from its mean. The squared deviation is adjusted to unit variance using the variance-covariance[6] matrix. A simple way to compute the MD is as follows:

- Compute the mean of each variable in the data
- Compute the standard deviation of each variable in the data
- For each record in the data
  - For each numeric variable on the record

[6] The variance-covariance matrix, which is similar to the correlation matrix (shown in Table 18), is a matrix displaying the covariance between each pair of variables. The diagonal of the matrix contains the covariance of each variable with itself, which is its variance.

    - Compute the difference between the value for the variable and the mean of the variable and divide by the standard deviation of the variable,
    - Square the result.
  - Sum the squared deviations of each variable for the record to derive the Mahalanobis depth

A computation using the algorithm above would ignore correlations between the variables that are accounted for in formula (2.3).

Using numeric variables in the automobile insurance data, a Mahalanobis depth was computed for each record. Since those records with a small value for this variable can be thought of as close to the center of the data and those with high values as on the perimeter of the data, the MD statistic can be used to "layer" the data. That is, the data can be ranked based on the MD value and grouped into quantiles. Table 7 displays the average MD statistic for data grouped into 20 quantiles.

Table 7

		Mahalanobis Depth
Quantiles of Mahalanobis Depth	1	*.78*
	2	*1.11*
	3	*1.35*
	4	*1.59*
	5	*1.83*
	6	*2.08*
	7	*2.33*
	8	*2.59*
	9	*2.89*
	10	*3.22*
	11	*3.61*
	12	*4.03*
	13	*4.59*
	14	*5.32*
	15	*6.41*
	16	*8.03*
	17	*9.52*
	18	*11.26*
	19	*13.31*
	20	*28.39*

Average Mahalanobis depth for 20 quantiles of the auto data

The analyst might choose to examine more carefully those records that are the furthest from the center of the data, i.e., those with the highest MD statistic. Table 8 presents a printout of 10 records, which were in the highest 1% of records on the MD statistic. Looking at the records in the table, the MD statistic seems to have identified records with unusual values on one or more variables. For the first, second, and fourth records, the number of drivers on the policy is six while the seventh record shows a negative value on the number of cars variable. The 6th record displays the year 490 as the license year while the last record shows a value of 2039 for this variable. This example indicates that the MD statistic has potential value for screening a large number of numeric variables for unusual values that may be data errors.

Table 8

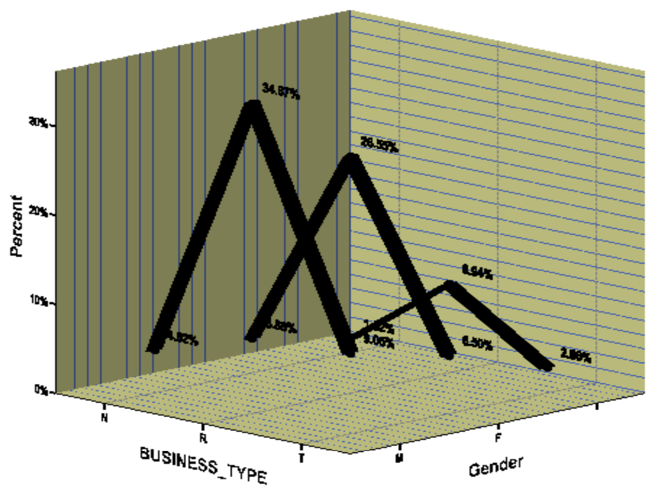
Policy ID	Mahalanobis Depth	Percentile of Mahalanobis	Age	License Year	Number of Cars	Number of Drivers	Model Year	Incurred Loss
22244	59	100	27	1997	3	6	1994	4,456
6159	60	100	22	2001	2	6	1993	0
22997	65	100	NA	NA	2	1	1954	0
5412	61	100	17	2003	3	6	1994	0
30577	72	100	43	1979	3	1	1952	0
28319	8,490	100	30	490	1	1	1987	0
27815	55	100	44	1976	-1	0	1959	0
16158	24	100	82	1938	1	1	1989	61,187
4908	25	100	56	1997	4	4	2003	35,697
28790	24	100	82	2039	1	1	1985	27,769

Listing of records with high Mahalanobis depth values

2.3 Categorical data: data cubes

The exploratory techniques described above can be applied only to numeric data. The techniques used to screen categorical data typically involve partitioning data. When exploring categorical data, the analyst typically uses data cubes, a topic that is covered in depth by Dasu and Johnson (Dasu and Johnson, 2003). Data cubes help us slice the data into chunks and see what is in the chunks. The data is partitioned into one-dimensional or multidimensional groupings. Frequency tables, cross tabulations and pivot tables are examples of data cubes. The partitions or cubes are then examined for unusual values.

Figure 14



Example of a data cube from auto data

Figure 14 illustrates a simple data cube. The figure displays the percentage of records in the data for each combination of business type and gender. In actual practice, the concept of data cubes is implemented by "slicing and dicing" the data into one-way or multi-way tabulation that reveal the structure of the data.

One of the most useful techniques for examining categorical variables is the one dimensional frequency table. Frequency tables list the values for the variables and the number of records containing the value. By reviewing such tables one can often detect impermissible values for the variable examined or learn other useful information about how the data is coded.

Tables 9 through 12 present the results of one-dimensional tabulations for some categorical variables in the data. Some observations can be made. There appear to be no data issues with the business type variable. However we note that about 14% of the records are missing a value for the gender variable. (Missing values will be addressed in more detail in section 2.4). We also note that marital status has the following codes: M, S, and D (which presumably denote married, single and divorced). In addition to these codes we find the codes '1', '2' and '4'. The coding of this variable appears to be inconsistent. Sometimes marital status is coded into a numeric code and sometimes it is coded into a character code. Since the analyst will want a consistent coding scheme, it will be necessary to contact the supplier of the data to learn the definition of the numeric codes.

Table 9

Business Type

	Frequency	**Percent**
N[7]	3607	10.2
R	25179	71.4
T	6498	18.4
Total	35284	100

Table 10

Gender

	Frequency	**Percent**
	5,054	14.3
F	13,032	36.9
M	17,198	48.7
Total	35,284	100

[7] N= New, R=Renewal, T=Targeted

Table 11

Marital Status

	Frequency	Percent
	5,053	14.3
1	2,043	5.8
2	9,657	27.4
4	2	0
D	4	0
M	2,971	8.4
S	15,554	44.1
Total	35,284	100

Reviewing the class code table below codes reveals that some class codes are very sparsely populated. It may be helpful to consolidate the data from sparsely populated cells into one "all other" category before conducting an analysis.

Table 12

Class Code

Code	Frequency	Percent
1	17,646	50
2	5	0
3	938	2.7
4	5,694	16.1
5	2,994	8.5
6	238	0.7
7	1	0
8	135	0.4
9	218	0.6
10	1	0
11	1,281	3.6
12	2	0
13	827	2.3
14	85	0.2
15	73	0.2
16	1,656	4.7
17	1,581	4.5
18	1,846	5.2
19	13	0
20	50	0.1
Total	35,284	100

Using macros or the command language for statistical software, the process of creating tabulations of the categorical variables can be automated.

2.4 Missing data

In large insurance databases, missing data is the rule rather than the exception. It is also not uncommon for some data to be missing on databases used for smaller analytical projects. Missing data complicates an analysis by reducing the number of records containing completely valid information that can be used. At a minimum, the uncertainty about parameter estimates will be increased, even when measures can be taken to adjust the data containing the missing values. It is not uncommon for the majority of records to be missing data on variables that are presumably in the database and available to the analyst. If a sufficient percentage of records on a given variable are missing a value, that variable may have to be discarded from the analysis. In some extreme circumstances, the missing data problem may be so severe that an analysis cannot be undertaken.

2.4.1 Detecting missing values

When an error is detected on a variable and its correct value cannot be determined, it is common to recode the value to missing. In addition, the original data may arrive with missing values on many variables. The analyst should screen each variable to be used in an analysis to determine the extent of the missing data problem. Most statistical software packages have default coding for missing values such as the period (.) or 'NA'. In addition, a coder may have used a specific value such as '99' as a code indicating a missing value. Missing data for character variables often takes the form of a blank field. Thus, it is necessary to completely understand the protocol for coding missing values that was used in assembling the data.

Tables 13 through 15 illustrate some of the issues that arise when screening for missing data. Table 13 shows the output of SPSS[8]'s frequency procedure and indicates that 5,042 records are missing a value for age and 5,034 records are missing a value for license year. The table also indicates that no records are missing for business type or gender. However, Tables 14 and 15, frequency tables of the values present for the business type and gender variables, indicate 14.3% of the records show a blank value for the gender variable, while all records contain one of the three legitimate values ('N', 'R' or 'T') for business type. In tabulating missing values for character data such as gender, it will be necessary to look at a listing of all possible values for the variable, and count those with a blank value as missing.

---

[8] SPSS is a vendor of statistical software. While the illustrations in this paper can be performed with free software such as R, the author found it convenient to use commercial software for some of the exploratory data analysis.

Table 13

		BUSINESS TYPE	Gender	Age	License Year
N	Valid	*35,284*	*35,284*	*30,242*	*30,250*
	Missing	*0*	*0*	*5,042*	*5,034*
Percentiles	25			*27.00*	*1,986.00*
	50			*35.00*	*1,996.00*
	75			*45.00*	*2,000.00*

Example of tabulation of missing values from statistical software

Table 14
BUSINESS TYPE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	N	*3,607*	*10.2*	*10.2*	*10.2*
	R	*25,179*	*71.4*	*71.4*	*81.6*
	T	*6,498*	*18.4*	*18.4*	*100.0*
	Total	*3,5284*	*100.0*	*100.0*	

Table 15
Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		*5,054*	*14.3*	*14.3*	*14.3*
	F	*13,032*	*36.9*	*36.9*	*51.3*
	M	*17,198*	*48.7*	*48.7*	*100.0*
	Total	*35,284*	*100.0*	*100.0*	

Table 16

Variable	Percent Missing
Age	14%
License year	14%
Number of vehicles	0%
Number of Drivers	0%
Marital status	14%
Territory	0%
Vehicle symbol	39%
Model Year	0%
Class code	0%
Business Type	0%
Policy type	0%
Number of claims	0%
Incurred losses	0%
Paid losses	0%
Paid allocated loss adjustment expenses	0%
Ultimate claims	0%
Ultimate losses and expenses	0%
Subrogation	0%
Earned premium	0%
Written premium	0%
Earned exposures	0%
Written exposures	0%
Zip Code	0%

Missing data percentages

Table 16 shows the missing value statistics for the variables in the data.

In addition to screening data supplied by others for missing values, the analyst needs to be alert to missing values he/she creates when performing calculations. Some functions, such as the log function will take on a missing value for some values supplied to it (in the case of the log function, most software codes the log of zero as a missing value). Most statistical software produces a log, which records the history of calculations completed and their results. Cody (Cody, 1999) recommends reviewing the logs of the statistical software the analyst is using for statements that missing values are being created as a result of transformations performed.

2.4.2 Types of missing values

The literature classifies missing data into three categories: 1) missing completely at random, 2) missing at random and 3) informative missing. (Allison 2002, Harrell, 2001). The category that missing data is assigned to has consequences for the strategies the analyst uses to address the missing data.

When data is missing completely at random for a variable, the fact that data is missing is completely independent of the values on any variables in the data. Under this assumption, a missing value on the age variable (which is missing in about 14% of the auto data) is unrelated to any potential dependent variables such as frequency of an accident or incurred loss ratio, as well as any potential independent variable such as territory, or class code. When data is missing at random, the probability of a missing value on a variable may be correlated with the values on other variables, but the value for the dependent variable is random after controlling for the other variables. For instance, if a value for age is more likely to be missing for single drivers, and the marital status is available on every record, an unbiased estimate of the age of a driver can be computed using age and marital status data from records where the age information is present. When data is informative missing on a variable, its true value is related to the value of the variable. Thus if age is systematically missing on very young drivers or very old drivers, the data is informative missing. This is also referred to as nonignorable non-response (Harrell, 2001).

2.4.3 Simple methods for missing values

One of the most common approaches to missing values is referred to as casewise or listwise deletion. This approach involves eliminating all records with missing values on any variable. Many statistical packages use this as the default solution to missing values. However, eliminating all records with missing values may result in discarding a large proportion of the data – data, which may contain valuable information that is useful to the analysis. For the auto data in this paper, in excess of 38% of the data would be discarded under this approach. Harrell points out that estimates based on casewise deletion of missing data are imprecise, biased or both. The imprecision results from the loss of a significant proportion of the data causing larger confidence intervals to apply to estimates based on the remaining data. Unless the data is missing completely at random the estimates will also be biased. If the value of a dependent variable such as claim frequency is, on average, higher or lower when the age variable is missing, a fitted model will be biased when all records missing values for age are deleted from the data. Table 17 indicates that frequencies in the auto data are lower on records missing a value for the age variable.

Table 17

		Claim Frequency
Age Missing	Missing	.04
	Present	.10
	Total	.09

Claim frequency vs. missing on age variable

Another approach that can be used for some statistical procedures such as linear regression is pairwise deletion of cases. For example, a linear regression can be estimated using only the means and covariances of the variables in the data. Each mean can be computed using all the records with values for the variable. The covariance between any two variables can be computed from all the records with a value for both variables. Pairwise deletion would eliminate only the records not containing the values on both variables from the computation of their covariances, but those records would be available for computing the covariances of other variables. That is, since each covariance is an estimate of how two particular variables co-vary, both variables must be present on a record for it to be used to compute their covariance. If data is present for those two variables but missing for a third, the record can still be used for part of the overall model estimation. Once the summary statistics have been computed, the regression parameters are estimated using these summary statistics. Allison (Allison, 2002) notes that pairwise deletion makes more use of the available data, therefore more efficient estimates (with smaller confidence intervals) are obtained when using this approach. However, Allison also notes that unless the data are missing completely at random, the estimates may be biased. Allison also points out that confidence intervals obtained using pairwise deletion are often under or overstated, depending on the rule used to determine the number of observations in the calculation of standard errors.

Another approach used to adjust data for missing values involves the use of dummy variables. A binary variable is created which is 0 if values are present for a given variable and 1 if values are missing. The dummy variable then becomes an independent variable in an analysis. Allison (Allison, 2002) points out that this approach is often biased. The method is equivalent to using the mean of the dependent variable for the missing values compared to the mean of data that do not contain missing values as a parameter estimate. When data are not completely missing at random, the result is likely to be a biased estimate.

2.4.4 Imputation

Imputation is a common alternative to the simple approaches listed above. It is used to "fill in" a value for the missing data using the other information in the database. A simple procedure for imputation is to replace the missing value with the mean or median of that variable. Another common procedure is to use simulation to replace the missing value with a value randomly drawn from the records having values for the variable.

Harrell points out that if a numeric predictor variable is independent of all other predictor variables, its mean or median can be substituted for the missing value (Harrell, 2001). It should be noted that the variability of the data will be understated, when a constant value is substituted for some of the missing values.

Since it is missing a value for a significant portion of the data, imputation will be illustrated using the age variable. The first step is to assess whether this variable is independent of the other predictor variables (in which case there would be no point in using them to estimate a value for age when it is missing).

A quick evaluation of the independence among numeric variables can be performed using the correlations between the variables. The correlation is a measure of the strength of a linear relationship between two variables[9]. Its value varies between -1 and 1. A correlation of zero indicates that a linear relationship does not exist between the variables. A correlation of 1 indicates a strong positive linear relationship between the variables and a correlation of -1 indicates a strong negative relationship between the variables. Most analytic software, including Microsoft Excel, have the capability of producing a correlation matrix. The matrix

[9] This correlation measure is sometimes referred to as the Pearson correlation.

displays the bivariate correlation of each pair of variables included in the correlation matrix request. The correlation procedure used for this paper also displays a test of the significance of the correlation. Table 18 displays a correlation matrix for selected numeric variables in the auto data. The table suggests that there is a relatively strong correlation between age and license year. There is a more modest correlation between age and model year. The test statistic indicates that both of these correlations are significant. The correlation measure used in this example only measures linear relationships and may miss or understate nonlinear dependencies between variables. It also does not provide a measure of dependencies between numeric variables and categorical variables or between categorical and categorical variables.

Table 18 – Correlation Matrix

		Age	*Drivers*	*License Year*	*ModelYear*	*No of Vehicles*
Age	Pearson Correlation	1.000	-0.005	-0.483	-0.056	0.006
	Sig. (2-tailed)	.	0.370	0.000	0.000	0.263
	N	30,242	30,242	30,226	30,237	30,242
Drivers	Pearson Correlation	-0.005	1.000	-0.027	0.061	0.235
	Sig. (2-tailed)	0.370	.	0.000	0.000	0.000
	N	30,242	35,284	30,250	35,279	35,284
License Year	Pearson Correlation	-0.483	-0.027	1.000	0.031	-0.009
	Sig. (2-tailed)	0.000	0.000	.	0.000	0.135
	N	30,226	30,250	30,250	30,245	30,250
ModelYear	Pearson Correlation	-0.056	0.061	0.031	1.000	-0.073
	Sig. (2-tailed)	0.000	0.000	0.000	.	0.000
	N	30,237	35,279	30,245	35,279	35,279
No of Vehicles	Pearson Correlation	0.006	0.235	-0.009	-0.073	1.000
	Sig. (2-tailed)	0.263	0.000	0.135	0.000	.
	N	30,242	35,284	30,250	35,279	35,284

The eta coefficient, η, is used to measure dependencies between numeric and categorical variables. It is typically used in conjunction with the analysis of variance (ANOVA) procedure, which is a common procedure for modeling a numeric dependent variable that has only categorical predictors (see Iversen and Norpoth, 1987). The formula for eta is:

$$\eta = \sqrt{\frac{SS_{between}}{SS_{total}}} \tag{2.4}$$

where SS denotes the sum of squared deviations[10]

As an example, Figure 15 indicates that there may be a dependency between age and the marital status variable. The eta coefficient measuring the correlation between age and marital status is 0.152. The F-statistic displayed with the output in Table 19 from an ANOVA indicates that the differences in age between categories of the marital status

[10] SS total is the total sum of the squared errors of the variable about its mean, while SS between is the sum of squared errors accounted for by the difference in mean valued between groups.

variable are statistically significant.

Figure 15

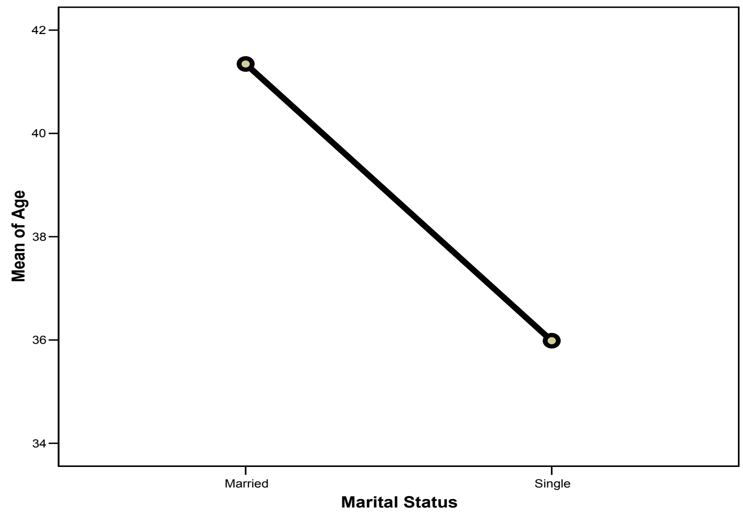


Table 19
ANOVA Output

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	*120.158,896*	*1*	*120.158,896*	*704,985*	*.000*
Within Groups	*5.150.923,322*	*30,221*	*170,442*		
Total	*5.271.082,218*	*30,222*			

Measures of Association

	Eta	Eta Squared
Age * Marital Status	*.151*	*.023*

ANOVA table showing test of statistical significance and correlation measure for age vs. marital status

Since the data indicate that age is correlated with other variables, using the other variables in the data for imputation of the missing values seems reasonable. One of the simplest procedures for imputation is linear regression. That is, regression is used to fit the model:

$$Age = a + b_1{*}X_1 + b_2{*}X_2 + \ldots + b_n X_n \qquad (2.5)$$

where X_1 through X_n are the other predictors in the data, including categorical variables[11].

Table 20 presents some output from the regression model[12]. Note that some variables, which are correlated with age (license year) could not be used because almost all records missing age are also missing license year. The predictor variables in the model are class code, coverage type, model year, number of vehicles and number of drivers. The regression had an R^2 of approximately 0.6, indicating that about 60% of the variance in age was explained by the model. Had the missing data been categorical (for instance if the analyst were imputing missing values for gender instead of age), logistic regression instead of linear regression could be used.[13]

[11] Categorical variables are included in the model through the use of dummy variables. A dummy variable is a binary variable that is either zero or one. Each value (minus one base category) of a categorical variable is a separate zero-one dummy variable. See Hardy (1993) for a more complete discussion of using dummy variables in regression. Most statistical software including that used for this paper automatically codes the dummy variables.

[12] A General Linear Model procedure was used to perform the analysis. This procedure is a generalization of linear regression and ANOVA.

[13] In logistic regression the dependent variable is binary. More advanced models using polytomous logistic regression are used when the dependent variable has more than two categories. See Hosmer and Lemshow (1989)

Table 20

Tests of Between-Subjects Effects Dependent Variable: Age						
		Type III Sum of Squares	df	Mean Square	F	Sig.
Source	Corrected Model	3,218,216	24	134,092	1,971.2	0.000
	Intercept	9,255	1	9,255	136.0	0.000
	ClassCode	3,198,903	18	177,717	2,612.4	0.000
	CoverageType	876	3	292	4.3	0.005
	ModelYear	7,245	1	7,245	106.5	0.000
	No of Vehicles	2,365	1	2,365	34.8	0.000
	No of drivers	3,261	1	3,261	47.9	0.000
	Error	2,055,243	30,212	68		
	Total	46,377,824	30,237			
	Corrected Total	5,273,459	30,236			

This illustration of imputation used a simple model to estimate missing values on a variable from the other variables in the data. A more complex method such as regression trees (Harrell, 2001, Gou, 2003) or MARS (Francis, 2003) could model complex structures in the data such as nonlinearities and interactions and might produce a more accurate estimate for the missing value. However, a detailed discussion of prediction methods is outside the scope of this paper.

Another approach for developing models when missing values are present uses the maximum likelihood method. The approach requires an assumption about the distribution of the data. For instance, the analyst might assume the data is from the multivariate normal distribution[14], and incorporate a specification for the missing data into the model. The estimation procedure finds the parameters that maximize the likelihood of the model given the data in the sample. Expectation maximization (Allison, 2002) is a common procedure based on the maximum likelihood approach that is used to estimate models in the presence of missing data. The maximum likelihood procedure will not be illustrated in this paper. An excellent introduction to the application of the EM approach to insurance problems is presented by Rempala and Derrig (Rempala and Derrig, 2003).

---

[14] Insurance data are typically positively skewed, as well as heavy tailed, so multivariate normality likely does not apply.

When the value of a variable is imputed, the statistics measuring confidence intervals for parameter estimates will typically be understated, because the "expected" value from a model is substituted for the missing value. This "expected" value will be missing a random component that is present in actual data when the values are present for the variable. Random imputation (Allison, 2002) addresses this concern by substituting the model's fitted value plus a random "error" term for the simple fitted value. The "error" term is typically simulated from a probability distribution that approximates to the distribution of the model's residuals, such as the normal distribution[15]. The new data with the imputed values will then have variability that more closely resembles the variability in data that do not have missing values.

2.5 The censorship problem: Using appropriate numeric data under censorship

Both the AAA standards of practice and the CAS and IDMA white paper on data quality cite appropriateness of the data as a key data quality concern. A common error relating to the appropriate use of insurance data results from ignoring censorship of insurance variables. Many insurance finance variables collected and used for analytical studies, which reside in insurance databases contain incomplete or censored values.

Insurance data is typically grouped into cohorts of similarly aged information based on when a policy covering an exposure is written (policy year) or when an incident giving rise to an accident occurred (accident year).[16] This means that as of any given point in time after the inception of an accident or policy period, only a portion of the final reported claims counts and paid loss amounts are known. This is a consequence of lags inherent in the reporting and settlement process for claims. Figure 15 on the next page, from the CAS Loss Reserve Seminar (Taylor, 2003) illustrates some of the lags affecting insurance data, which cause insurance data to be incomplete. That is, some claims are not reported for a number of weeks and in some cases a number of years after the incident causing the claim occurred. While most personal automobile insurance claims are reported within a year of their occurrence, there are some lines of business, such as professional and products liability,

[15] For many statistical models, errors are assumed to be from a normal distribution, but other distributions are likely to be more appropriate for insurance variables. Bootstrapping residuals from the sample of actual residuals is a distribution free way to randomly generate the residual term in random imputation.
[16] Data can also be grouped according to other date variables, such as when the claim was reported.

where decades may pass before a claim is reported. There are additional lags in the investigation and settlement of claims. Claims that are litigated might take many years to reach their ultimate or final settlement value. When analyzing data grouped by policy year, there are additional lags because policies usually are sold throughout the year, and for policies sold in December, accidents may occur as late as December of the following year

Figure 16, (Taylor, 2003) illustrates how financial values evolve over time for a hypothetical sample of insurance data. The figure illustrates how it can take many years for the final settlement values for all the claims in a given accident year to be known. Until the year is very mature and all claims are settled, the analyst must work with incomplete, or censored, data and make appropriate adjustments. Figure 17 illustrates the development over time of cohorts of paid losses organized by accident year. The more recent the accident year, the more immature the paid loss data is and the less that is known about the "ultimate" or final settlement value of the claims.

Figure 16

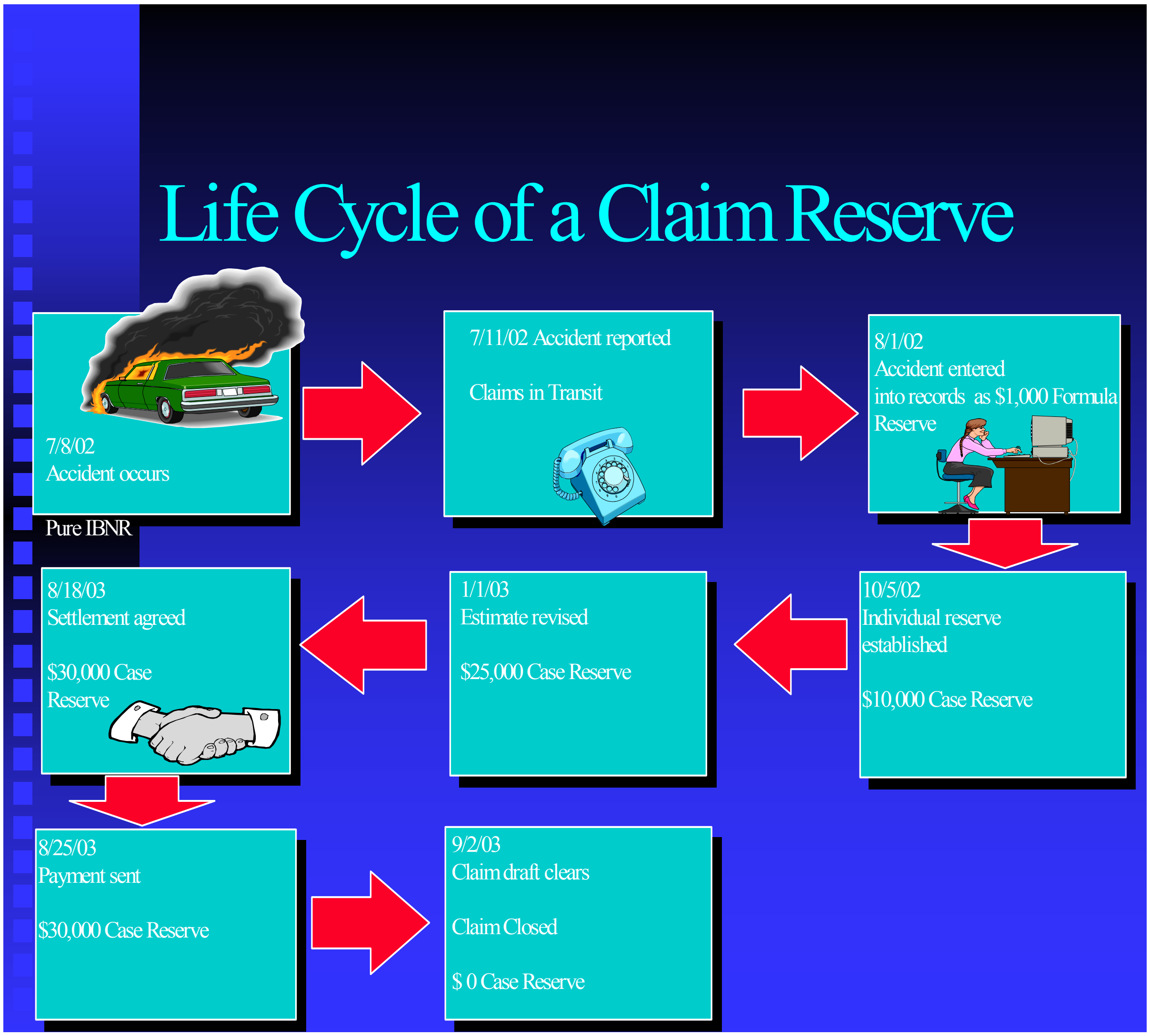
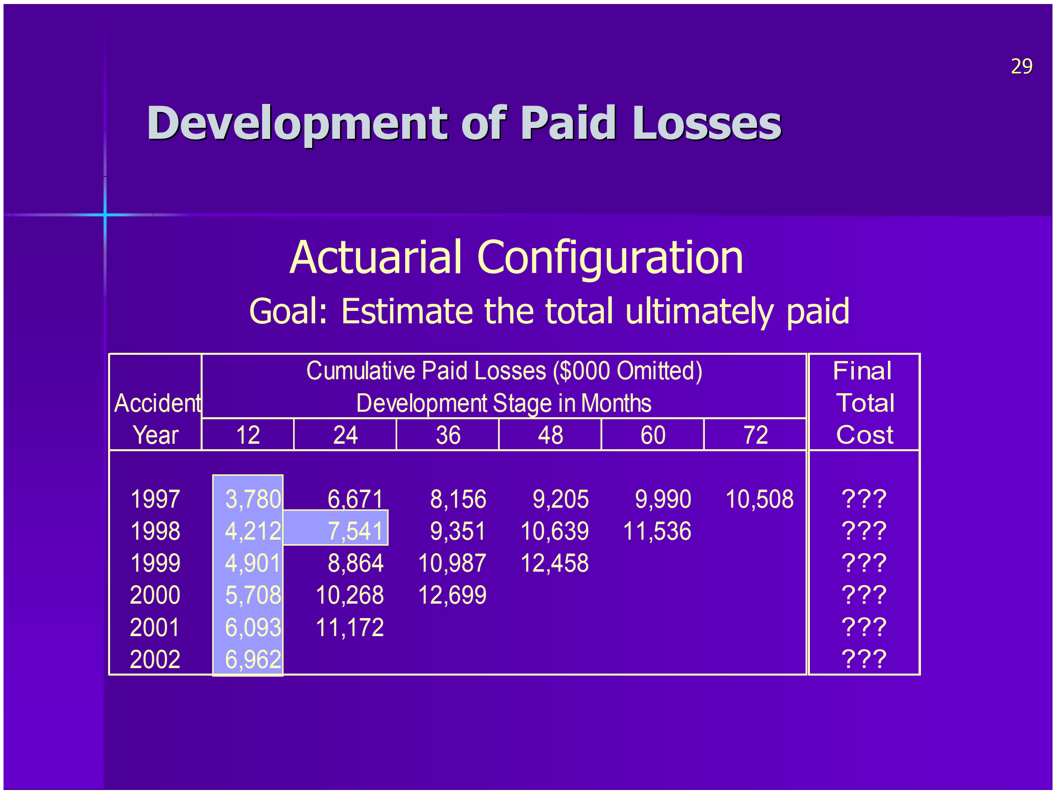

Life Cycle of a Claim Reserve
7/8/02
Accident occurs
Pure IBNR
7/11/02 Accident reported
Claims in Transit
8/1/02
Accident entered
into records as $1,000 Formula
Reserve
10/5/02
Individual reserve
established
$10,000 Case Reserve
1/1/03
Estimate revised
$25,000 Case Reserve
8/18/03
Settlement agreed
$30,000 Case
Reserve
8/25/03
Payment sent
$30,000 Case Reserve
9/2/03
Claim draft clears
Claim Closed
$ 0 Case Reserve

Figure 17



Accident Year	Cumulative Paid Losses ($000 Omitted) Development Stage in Months 12	24	36	48	60	72	Final Total Cost
1997	3,780	6,671	8,156	9,205	9,990	10,508	???
1998	4,212	7,541	9,351	10,639	11,536		???
1999	4,901	8,864	10,987	12,458			???
2000	5,708	10,268	12,699				???
2001	6,093	11,172					???
2002	6,962						???

While casualty actuaries are familiar with the impact of performing analyses with data that is not fully mature, other users of insurance data all too often are not. A concrete example, which the author actually observed, involved a benchmarking analysis. Benchmarking generally involves the comparison of one entity's performance against that of a standard or base. The standard is often "the industry', or other companies selling the same product. Sometimes it is a selected group of competitors. The example of inappropriate use of property and casualty insurance data involves a benchmarking tool that was sold to insurance companies, insurance brokers and third party administrators in the mid 1990's. The purpose of the tool was to enable a company to compare the average severity of its settled claims with those of its competitors. A typical user would compare calendar period (typically calendar year) closed claim severities for a given company with the average calendar year severities of all the other companies in the database. When the analyst was benchmarking the claims of a new company or program, it was easy to "prove" that the program was better than the industry, as their claims data consisted only of immature claims whose average severities would be considerably lower than an industry portfolio consisting of a more seasoned mixed of claims whose average maturities, and therefore average severities would be higher.

Tables 21 and 22 help illustrate how the censorship problem affects calendar year comparisons. Table 21 displays a hypothetical distribution of claims settlement by development age. In this table, assume that the development age denotes the length of time in years since an accident occurred. Development age 1 refers to all claims that settle within one year of the occurrence of an accident. Column 2 of the table shows the average closed claim severity for all claims, which settled at a given age. Table 22 displays the effect of comparing a company that has been in business for only one year to an industry benchmark composed of companies, which have been in business for many years. For simplicity, we assume no impact from inflation in the illustration. The new company, because its claims inventory is immature and is composed only of the claims settled quickly for modest amounts, has an average severity that appears to be much better than the industry, even though its claims settle for exactly the same amount as similarly aged industry claims.

Table 21

Age (Years)	Closed Claim Severity	Percent of Claims
1	500	25%
2	1,000	50%
3	5,000	15%
4	10,000	10%

Distribution of claims and average settlement amounts by age.

Table 22

New Company			
Accident Year	**Age**	**Severity**	**Percent**
2003	1	500	100%
Average Severity		500	

Industry			
Accident Year	**Age**	**Severity**	**Percent**
2003	1	500	25%
2002	2	1,000	50%
2001	3	5,000	15%
2000	4	10,000	10%
Average Severity		2,375	

Illustration of a naïve comparison of a new company or program to a mature industry sample.

Several strategies are available to address the problem of censorship in insurance data. The first strategy is to sample only records with the same "as of dates", i.e., use similarly aged data. That is, the study data in the example above might consist only of claims with a settlement age of one year. A drawback of this approach is that only a portion of the sample will make it into the study and these claims may not be representative of the values that would be observed on a more mature body of data. If only mature claims are used in the study, important patterns occurring only in recent data may not be detected.

The second alternative is to adjust all values to an ultimate basis[17], using a standard actuarial procedure such as development. Using this approach, an unbiased estimate of summary statistics, such as average ultimate severities or ultimate loss ratios will be obtained when comparing one group to another from the data. A drawback of this approach when it

[17] Ultimate values are actuarial estimates of the final settlement value

is applied to incurred losses (that is paid losses plus case reserves) is that each claim in a database is assumed to develop by the same percentage, when in fact some claims will develop by a much larger percentage than others and some may develop downwards. When applied to closed claims, an amount other than the real settlement value of the claims is obtained, since the final settlement of the claims is presumably known. This is the approach used with the auto data that serves as an example in this paper. Ultimate values for reported claims and incurred losses were estimated and incorporated into the database for use in analytical studies.

Another alternative is to weight records or adjust the records to a constant mixture by age when computing averages or other statistics. That is, when the mix of claims by age varies across groups, the analyst could analyze data on an apples-to-apples basis by applying the same weight to all claims of a given age. Continuing the benchmarking example, Table 23 shows how we might mix adjust our claims settlement data before computing average claim severities. If claims are mix-adjusted using the mix of claims by age for the new company, the resulting mix-adjusted severity is the same for both the new company and the industry and an appropriate comparison is made. If however the mix of claims by age for the industry is used to compute the averages, a misleading result is obtained because zeros are incorporated into the average for the new company, for severities of older aged claims. Thus, the analyst must be careful in selecting weights and applying the mix adjustment.

Table 23

Using New Company Mix			
Age	**Weight**	**New Company**	**Industry**
1	100%	500	500
2	0%	0	1,000
3	0%	0	5,000
4	0%	0	10,000
Average		500	500

Using Industry Mix			
Age	**Weight**	**New Company**	**Industry**
1	25%	500	500
2	50%		1,000
3	15%		5,000
4	10%		10,000
Average		125	2,375

Sampling techniques are also sometimes used to address the censorship problem. Returning to the data from the personal automobile example, Table 24 presents statistics on the percentage of all policyholders reporting a claim. The data in this example are valued approximately midway through 2003. It can be seen that policy year 2003 displays a much lower percentage of records with claims than the prior policy years. The claim frequency is roughly one-fourth the rate for the other years. The low frequency must be assumed to result from censorship of the data as the 2003 policy year was not a complete year at the time the sample was created. If we were analyzing the data to find predictors of claim frequency, we might choose to over sample records with claims compared to records with no claims for the 2003 year. That is, the analyst might sample records for 2003 that have claims, at 4 times the rate of records that do not have claims. In addition, the 2002 policy year will not be a complete policy year until December of 2003, so we may wish to over sample records with claims for the 2002 year also.

Table 24

Policy Year	Percent with Claim
2000	8.10%
2001	10.20%
2002	7.50%
2003	2.00%
Total	7.40%

2.5 Metadata: What is in the data?

Metadata is a term used by data management and data quality professionals to denote data that describes the data, i.e. the documentation of the contents of a database. This would include a listing of all fields in the data, along with a description of what is contained in each field. The metadata will likely contain a list of variables or field names. Each field listed should be defined clearly and the data that is in the field described. Thus, in the metadata, the field pol_eff_date is defined to contain the policy effective date and should contain only date values. The permissible ranges of the values (i.e. 1/1/2000 through 6/30/2003 on the policy effective date field) should be specified. Metadata should also define the labels in categorical data. As an example, recall that (see Table 11) six values are present in the data for the marital status variable. Table 25 displays one scenario for defining the values in the marital status field.

The definition of values such as paid and incurred loss should specify whether legal and other claim adjustment amounts are included and whether the data in the field is net or gross as to subrogation and recoveries. Incurred loss metadata should also specify whether the incurred losses represent an estimate of ultimate incurred losses or whether the amounts represent paid losses and case reserves as of a given valuation date. If the latter, the valuation date should be specified.

Table 25

Marital Status Value	Description
1	Married, data from source 1
2	Single, data from source 1
4	Divorced, data from source 1
D	Divorced, data from source 2
M	Married, data from source 2
S	Single, data from source 2
Blank	Marital status is missing

Description of marital status field

The more complete and comprehensive the metadata, the better. A complete description of the contents of a database is important to the appropriate use of the data. Good metadata can assist the analyst in avoiding misunderstandings that result in revisions of the analysis when the contents of a variable are discovered to be other than what it was assumed to be.

One problem that occurs frequently when comprehensive documentation is not maintained is that the person(s) familiar with the contents of a database leave a company and no one is left who is familiar with some of the quirks of the data. Hence, maintenance of adequate documentation describing data can help avoid problems associated with relying exclusively on people's memories of what is contained in the data.

Olson (Olson, 2003) points out that one output of a data screening process should be additional metadata. That is, when data is screened the analyst does not actually begin with complete metadata, including a description of data anomalies and a detailing of fields with missing values. Once the data is screened, new metadata should be created describing the structure of the data, including what was learned during the data screening process.

3. DISCUSSION AND CONCLUSIONS

The problem of "dirty" data is ubiquitous. Data often contain erroneous values and must be scrubbed to remove such values. Data often are incomplete with values missing on many of the variables that are of interest to the analyst. If values for the missing data cannot be supplied, the data needs to be adjusted for the missing values.

This paper presented a number of methods, which can be used to screen data for unusual values. Many of the methods presented are graphical and have been in the statistical literature for many years but are not widely used by actuaries. These include histograms and box and whisker plots. When applying these procedures to insurance data, adjustments to the procedures such as filtering selected values and graphing on a log scale are sometimes needed in order to obtain useful results. This paper has also presented a more recent approach to screening data: data spheres. The MD statistic based on the data spheres concept can be used to screen numeric variables simultaneously for unusual values. Once an unusual value or outlier is detected, the analyst can determine whether the value represents an error, or whether it can remain in the data for use in an analysis.[18]

This paper also discussed the missing value problem and presented several methods, which can be used to adjust for the missing values when performing an analysis. The imputation approach was introduced and a simple implementation of imputation was illustrated. More advanced procedures for doing data imputation are found in Allison and Harrell (Allison, 2002, Harrell, 2003). This paper addressed the inappropriate use of censored data. Censored data occurs frequently in property and casualty insurance databases. The paper suggested approaches, which can be implemented in the presence of censorship.

The importance of good metadata was also discussed. The data analyst ideally will be supplied a comprehensive description of the data in a database. Having a good

[18] While extreme values occur in insurance data, even when the value is legitimate, the analyst may want to take measures to reduce the influence of the value on estimates. Robust methods and other procedures, which are resistant to outliers, can be applied under such circumstances, but these methods are outside the scope of this paper.

understanding of the data can help to avoid costly problems.

A topic that was not discussed in this paper, but which is addressed in some of the data quality literature is measuring the quality of data. Dasu and Johnson (Dasu and Johnson, 2003) present rules for evaluating the quality of a database. An objective of such a measure is to provide feedback to data managers, which will assist them in the improvement of the quality of their data. Thus, many of the data quality authors urge users of data to become effective advocates of improvements to data quality. However, even with efforts to improve the quality of data, data quality problems are likely to continue to exist. In Dasu and Johnson's words "In the end, the best defense is relentless monitoring of data and metadata".[19]

[19] Dasu and Johnson, p188.

Acknowledgment

The author acknowledges the editorial assistance of Patricia Francis-Lyon, Michael Francis, Virginia Lambert and Jane Taylor.

4 REFERENCES

[1] Actuarial Standards Board, Actuarial Standard of Practice No. 23, Data Quality, July 1993
[2] Allison, P., *Missing Data,* Sage Publications, 2002
[3] CAS Committee on Management, Data and Information, "White Paper on Data Quality", *Casualty Actuarial Society Forum*, Winter 1997, pp145 – 168
[4] Casualty Loss Reserve Seminar Planning Committee,. "Basic Track I", *Casualty Loss Reserve Seminar*, 2003
[5] Carter, N. and Bradley, S., "Date of Reference for Age and Birth Date used by Respondents of Census 2000", Census 2000 Evaluation H.10, November, 2002
[6] Chambers, J, Cleveland, W., Kleiner, B. and Tukey, P., *Grapical Methods for Data Analysis*, Wadsworth International Group, 1983
[7] Chen, P and Popovich, P, *Correlation*, *Parametric and Nonparametric Measures*, Sage Publications, 2002
[8] Cleveland, W., *Visualizing Data*, Hobart Press, 1993
[9] Cody, R., *Cody's Data Cleaning Techniques Using the SAS Software*, SAS Institute, 1999
[10] Dasu, T and Johnson, T, *Exploratory Data Mining and Data Cleaning*, Wiley, 2003
[11] Fox, J, *An R and S-PLUS Companion to Applied Regression*, Sage Publications, 2002
[12] Francis, L., "Martian Chronicles: Is MARS Better than Neural Networks?", *Casualty Actuarial Society Forum*, Spring, 2003, pp. 75 – 102
[13] Gou, L., "Applying Data Mining Techniques in Property/Casualty Insurance", *Casualty Actuarial Society Forum*, Winter, 2003, pp. 1 – 26
[14] Hardy, M. *Regression With Dummy Variables*, Sage Publications, 1993
[15] Hartwig, F. and B. Dearing, *Exploratory Data Analysis*, Sage Publications, 1979
[16] Harrell, F., *Regression Modeling Strategies*, Springer, 2001
[17] Heitjan, J and Reboussin, D., "A Model for Coarse Grouping with Application to Survey Data", The Rand Corporation, 90024-1766
[18] Hosmer, D and Lemshow, S., *Applied Logistic Regression*, Wiley, 1989
[19] Iversen, G and Norpoth, H, *Analysis of Variance*, Sage Publications, 1987
[20] Jacoby, W., *Statistical Graphics for Univariate and Bivariate Data*, Sage Publications, 1997
[21] Popelyukhin, A., "Let me See: Visualizing Actuarial Information", *Casualty Actuarial Society Forum*, Winter, 2001, pp. 399-426
[22] Olson, J., *Data Quality: The Accuracy Dimension*, Morgan Kauffman Publishers, 2003
[23] Rempala, G. and Derrig, R., "Modeling Hidden Exposures in a Claim Severity Using the EM Algorithm", *Casualty Actuarial Society Forum*, Winter, 2003, pp. 75 - 102
[24] Silverman, B. W., *Density Estimation*, Chapman and Hall, 1986
[25] Taylor, J.,. "Basic Track I", *Casualty Loss Reserve Seminar*, 2003
[26] Tufte, R., *Visual Displays and Statistical Thinking: Displays of Evidence for Making Decisions*, Graphic Press LLC., 1997
[27] Venibles, W and Ripley, B, *Modern Applied Statistics with S-PLUS*, Springer, 1999

Abbreviations and notations

AAA, American academy of actuaries
CAS, Casualty Actuarial Society
CLRS, Casualty loss reserve seminar
IDMA, Insurance Data Management Association
MD, Mahalanobis depth

Biography of the Author

Louise A. Francis is a Consulting Principal at Francis Analytics and Actuarial Data Mining, Inc. She is involved in data mining projects as well as conventional actuarial analyses. She has a BA degree from William Smith College and an MS in Health Sciences from SUNY at Stony Brook. She is a Fellow of the CAS and a Member of the American Academy of Actuaries. She is chair of the CAS committee on the theory of risk, and is a frequent presenter at industry symposia. She published two previous papers in the Data Management, Quality and Technology Call Paper Program: "Neural Networks Demystified" (2001) and "Martian Chronicles: Is MARS Better than Neural Networks" (2003).

The Games We Play: the Future of DFA Models' Interfaces

Aleksey S. Popelyukhin, Ph.D.

Abstract

Motivation. (Call for Papers). The author *strongly* believes that future generations of DFA software should employ *cardinally different* interfaces in order to reflect growing *complexity* and provide necessary *flexibility* for the models.

Method. Work by *analogy*. Among existing software products the author found one with the interface almost ideally fitting to the future needs of DFA packages.

Results. *Shocking* (but only at first): the software product with the "ideal" DFA interface is neither analytical nor calculational package, but a game, a computer simulation game. Indeed, computing gaming industry is the most creative and innovative niche of software development, where computer-human interfaces are given *the highest* priority. It is only *logical* to look there for the ideas for the best interface, especially for such *interactively challenged* products as DFA models.

Conclusions. If the author is correct in predicting the communicative requirements of the future DFA models, then DFA developers should look closely at some computer games and borrow ideas for interface design.

Keywords. Actuarial software design.

1 INTRODUCTION

The abbreviation DFA means Dynamic Financial Analysis. But, if you look closely, it is not that *dynamic*. Current DFA products accept some parameters from the user and using predefined algorithms (called model) launch a Monte-Carlo simulations to calculate distributions and some statistics of predefined variables at predefined point(s) in time. It is indeed a great achievement to have working DFA models, but it is *definitely* not the end of the road. Indeed, after user's initial input, the model is *closed for interactions*. Essentially, user is allowed to make decisions, that is, to choose options available to him (parameters) or to define a strategy (model algorithms), only once. There is nothing *dynamic* about it. In real life decisions are made *constantly* as a reaction to the changing environment. For DFA it means that some simulation passes that contribute to the final statistics would have no chance to exist or arrive to different value if user had a change to react during the run. That is, some investments could be sold, some reinsurance could be canceled, some capital reallocated, but only on some passes and not on the others. So, for DFA system to mimic decision making process better, an interface has to be built to allow user monitor simulation passes and interact with them in "real time" and/or modify strategy "on the fly". While these capabilities are unthinkable in Excel/@Risk paradigm, they are quite commonplace in the …

computer gaming universe. Let us show an astonishing *analogy* between an imaginary DFA system and an existing simulation game.

2 ANALOGY: LET THE GAME BEGIN

2.1 Loading…

Historically Personal Computers had two major applications ("killer apps"): VisiCalc and Tetris. People were buying PCs just to run a spreadsheet or a game, in other words, a powerful "what-if" analysis tool and a visually appealing entertainment program. Since the early 1980s both categories improved dramatically: Nowadays spreadsheets (extended by internal programming languages and connected to external databases) are housing quite sophisticated models, while games (enhanced by impressive 3-D graphics and intuitive interfaces) are featuring believably immersive environments. Nevertheless, both categories remain the main reason for buying a computer: their utility is still unsurpassed by other types of applications.

Now imagine an application that combines the visual appeal and intuitive interface of a game with the analytical power of a spreadsheet. It may be achieved either by adding visual interfaces to a spreadsheet or by adding analytical calculations to a game. The latter, apparently, seems more realistic: it is possible to find an existing game (or a genre) that may serve as a visual shell for the existing dynamic risk models.

2.2 Visual Metaphor

Dynamic Risk Modeling (in one way or another) deals with the random processes, i.e., studies values changing in time. It attempts to reflect numerous economic aspects in the life of an insurance company. And the more sophisticated the model is, the more complicated it is for the user to grasp how it functions and even less so interact with it.

These models usually simulate the growth of losses, flow of investments, changes in pricing conditions and consequences of catastrophes, all the while trying to properly deal with the time component as well as with geographical dissemination of the risk…

Amazing, but that is **exactly** the subject of the numerous city/empire/railroad simulation games. Indeed, these games visualize the growth of the buildings, flow of funds, changes in trade conditions and consequences of disasters. To think about it, 'growing buildings' may be interpreted as losses, different 'zoning areas' as different lines of business, 'bulldozing cost' as brokerage fees and 'earthquakes' as (evidently) earthquakes.

Once one makes a mental substitution (re-labeling), one should realize he has a tremendously capable visually rich and ready for utilization interface for his risk model. But even more important than visualization is the fact that such an interface is **interactive**.

2.3 Functional Metaphor

Let's continue our "risk model as a city simulation" analogy to the functional level.

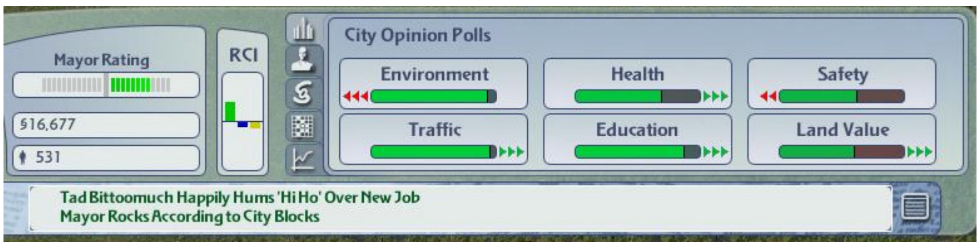
2.3.1 Mayor

User is a Mayor, an Emperor or a Tycoon. He has advisors – AI constructs each covering its area of expertise: expenses, revenues, services, market conditions. Mayor observes life of the city/empire, consults with advisors, makes decisions and **acts** in extraordinary situations. That is very similar to the role of the CEO who has the advice of his Actuary, Accountant and Claim Adjuster at his disposal. The goal of the game to maintain the financial health of the city/empire and the Mayor has quite a few instruments to achieve that.

2.3.2 Monitoring

The quintessential component of correct decision making is accurate and timely information. "Game-like" interfaces provide quite an extensive collection of monitoring tools. The variety of indicators available for observation is stunning. Along with numerous statistics and indicators, "game-like" interfaces usually feature distribution histograms, geographical maps, color-coded diagrams and charts.

Most important, indicators are displayed prominently on the screen so the user may observe the changes in their values in "real" time. To be even more useful "game-like" interfaces display the indicator's values along with their *derivatives*, that is, their direction and speed of change.



In the event when an indicator or some combination of them reaches a benchmarked value an alert is generated. The whole purpose of monitoring mechanisms and alerts is to prompt the user to make decisions and interact with the system.

2.3.3 Making decisions

One should agree that the ability to modify a simulated scenario "on the fly" makes the model truly dynamic. In the gaming paradigm the user may affect the system in different ways. The most straightforward one is to modify values of the system's parameters. A Mayor in the game may cut expenses or borrow money. He may decide to bulldoze some areas which can be interpreted as a commutation of some treaties, or he may cut supplies to an unprofitable region which would represent a switch to run-off mode.

Immediate feedback and rich visual metaphor in a "game-like" environment should help one to achieve a better grasp on the consequences of the user's actions.

2.3.4 Actions

In the game, the user's actions are not limited to setting some parameters values. The user may place structures that have local effect or design infrastructure (like electrical subsystem) that affects whole regions. In this interpretation placing a 'police station' that reduces the crime rate corresponds to hiring a "bill monitoring" firm to reduce medical expenses or replacing a team of lawyers in order to decrease legal fees.

It appears that any action that the management of the real company can possibly make has an acceptable analog in the game paradigm. It means that management's actions can be incorporated into a model as well.

2.4 Extras

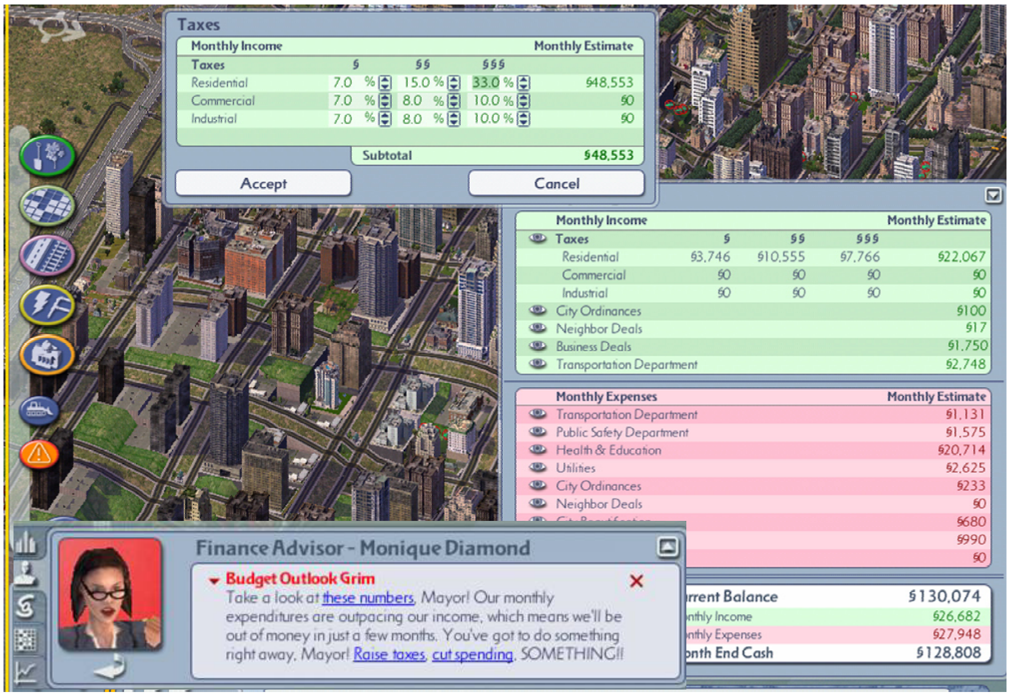
Modern games of the "city/empire simulation" genre implement many useful aspects that are seamlessly incorporated into their interface. Designers of the risk models who decide to utilize a "game-like" interface may consider these features as a free bonus.

2.4.1 Catastrophes, geography

In order to look as life-like as possible games simulate natural disasters. Not only do they incorporate visual representations of hurricanes, foods, earthquakes and tornados, but they represent damage in a geographically accurate manner. In fact, the object-oriented nature of these games allows the user to extend them and import unique real life structures, cities and even whole regions.

2.4.2 Financials

The games of this genre usually have a module responsible for simulating economic conditions. Their interface presents pro-forma like accounting reports and an ability to manipulate with economic parameters. In fact, these games have become so sophisticated in monitoring financials that they have even implemented AI advisors in this area.



2.4.3 Traffic

Some games feature quite advanced traffic simulators. The traffic patterns in these games take into account the time of the day, seasons, road capacities and even simulated routes from residential to industrial and commercial areas. That is a little bit more sophisticated than just finding a factor in a rating table.

2.5 Initial Setup

2.5.1 Terra-form

A very rich visual environment of the game requires a very powerful mechanism for its initial setup. Modern games have a convenient interface even for setting up starting values of the model parameters, distributions and maps. It is implemented as just an additional mode; game designers call it (quite suitably) "God mode". Being "God" the user can specify initial values for all economic parameters, outlay model settings, and define distributions, that is, (in a "game-like" metaphor) "terra-form" business landscape.

3 BENEFITS: NOT A GAME ANYMORE

Evidently, games provide a richer visual environment for the risk modeling, but this is only a small part of the story. Switching to a "game-like" interface complemented by improved modeling approaches may bring numerous benefits.

3.1 Training

3.1.1 Thinking with images

The way we think about concepts and operate with them heavily depends on the method we use to represent them. Rules of manipulation with the string of algebraic symbols significantly differ from the rules of manipulation with geometric shapes and curves. In one environment we may look for the values x that makes an integral such as $\int_x^\infty \frac{\beta}{\theta(1+\beta)^2} \exp\left(-\frac{x}{\theta(1+\beta)}\right) dx$ smaller than 5%, in another we may look for the vertical lines that make an area under the (distribution density) curve smaller than 5%. In one paradigm we may talk about first and second derivatives, in another we may discuss growth and convexity.

The "game-like" interface adds yet another form of imagery and various ways to manipulate it, which are predominantly more intuitive and convenient than formulas and charts.

3.1.2 Gaining decision making experience

Given that behind the glamorous interface lies a decent simulation engine, a risk model may serve as a management training tool. By visualizing consequences of every decision such an environment may help to polish the management style of key decision makers. It may also serve as a test-bed for new strategies and tactical innovations.

3.1.3 Modeling disastrous or unusual events

"Game-like" interfaces may prove to be an ideal playground for so-called stress testing. It is much more useful to study scenarios that include natural disasters or macroeconomic shifts in an **interactive** environment, performing actions exactly when (and where) they are needed most.

3.2 Fine-Tuning

It is conceivable that "game-like" interfaces will not replace, but rather assist in enhancing existing risk models. They can be used to fine-tune some aspects and design decisions of existing DRMs.

3.2.1 Choosing criteria

Risk models oftentimes rely on a set of benchmarks and criteria that seem to be chosen more or less arbitrarily. It would be very educational to **see** what happens if, for example, confidence intervals are shrunken or probability-to-ruin is replaced by some other evaluation criteria. Designers may drastically improve a model's relevance just by observing what combination of indicators triggers a user's action.

3.2.2 Refining strategies definitions

By allowing users to perform a multitude of actions model designers may refine the list of available strategies. They may build into the model automatic responses in order to improve the validity of scenarios. Without such corrective mechanisms some scenarios may never happen. In essence, designers have to "teach" their models When (and Where) to do What.

3.2.3 Business Processes

Observing people "playing" with "game-like" models may help to identify sequences of actions in different situations: in other words, management style. It may also help to analyze business processes and the chain of command to pinpoint problems and deficiencies.

3.2.4 Assumptions Testing

A "game-like" environment may also serve as a testing ground for numerous assumptions that are incorporated into the model. Incorrect assumptions and improbable parameter values may produce improbable situations that could be easier to spot in a visual environment.

3.3 Self-Education

Models with "game-like" interfaces may record and analyze a user's responses in order to use them later in a run of simulated scenarios. Oftentimes it is difficult for an expert to explain his actions in a formalized manner suitable for modeling. By watching him play designers or the software itself may try to determine rules that govern the user's behavior. In other words, "game-like" interfaces may assist in an attempt to build a self-educating model, that is, "evolving AI" system.

3.4 Monitoring

Interactivity (one of the main benefits of the "game-like" interface) is useless without parameter monitoring. Indeed, interaction has to be **triggered** by events or alerts that are either text-based or visual.

3.4.1 Multiple factors at once

"Game-like" interfaces provide an ability to choose which indicators are to be permanently monitored. User may observe the behavior of a single or multiple parameters, scalars or maps, charts or diagrams, or all at once, in "real time".

3.4.2 "Real time"

One of the most important features of "game-like" interfaces is animation. User may observe how a particular scenario runs as time passes. Surely, the user is given full control over the timeline: he can speed things up or slow the animation down or even pause for detailed investigations and a thorough analysis. Observing the system in a dynamic setting, though, adds a whole another dimension to one's perception and comprehension of the system. If a picture is worth a thousand words, than an animation is worth a thousand pictures!

3.4.3 Visual cues

In some situations diagrams and maps work much better than numbers and text. Two-dimensional maps, for example, much better represent such notions as "concentration" or "dissemination" than any singular value or sentence.

In addition, "game-like" interfaces may utilize visual hints to attract user's attention: fire icons, blinking symbols or color change may do the trick.

3.5 Dynamic Response

3.5.1 Time is visualized

The capacity to comprehend is greatly improved by using animation to represent time. Unlike traditional DRMs, "game-like" interface provides direct access to a timeline for dynamic responses.

3.5.2 Non-linear strategies

"Game-like" interfaces provide an environment in which responses and their timing are completely flexible and, presumably, more suitable to each particular situation. No preset strategy, no matter how sophisticated, can match the effectiveness of the dynamic response system.

3.5.3 Feedback

Feedback is used to represent controlling mechanisms as well as the forces that affect dynamic processes. Many economic indicators in an insurance company have a feedback: changes in an indicator's value triggers external actions that in turn change the indicator's value. Essentially, an indicator's "behavior" in time may be dissimilar for different values of the indicator. Analytically speaking $u'(t) = f(t,u)$, that is, an indicator's derivative is a function not only of time, but of the indicator's value as well.

3.5.3.1 Reserves

Once the reserves grow "too large" (ultimate expected loss value exceeds some benchmark) and the pressure from rating agencies becomes unbearable some companies may start writing commutations or switching to run-off mode, thus altering payout patterns and, consequently, changing ultimate expected loss value.

3.5.3.2 Prices

Pricing is also subject to feedback. High premiums may result in larger profits attracting competition which places downward controlling pressure on prices.

3.5.3.3 Investments

Badly performing investment instruments may get reinvested altering in turn their return rate. Another famous macroeconomic example of a system with feedback is inflation. The controlling mechanism in this case is the Federal Reserve Board.

While feedback as a non-homogeneous effect is almost impossible to implement as a closed analytical model, it is usually not such a big obstacle in the design of simulations. Sometimes simulation is the only way to model feedback making a "game-like" interface a natural environment for effective representation of controlling effects.

3.6 Investigations

To quote designers and promoters of the Public Access DFA Model: "In examining the DFA runs, many questions were raised [by managers of the real company] about what might have been causing adverse experience. It was suggested that the program be revised to capture detailed financial data on any simulation where surplus fell below a certain level. Thus, the managers could look at what caused the problems in order to better avoid them". Apparently, an ability to run specific scenarios (in order to identify circumstances causing unacceptable performance) is of great value for decision makers.

3.7 Presentation of Results

One of the main purposes of the "game-like" interfaces is to incorporate the presentation of the results in order to avoid complicated explanations in the end. Results are "built-in" in the interface. They are by-products of the values that were monitored. To the user, who had a chance to observe the process, the results are self-evident. Not only does he get a snapshot of the various monitored indicators, he also gets an idea of the *direction* in which they were moving and their *behavior*. It is one thing to merely observe that an indicator has reached the value of A, it is yet another to learn that it actually "dropped to the value A" or "seesawed to the value A swinging back and forth". The latter feedback is, obviously, much more informative. Results themselves could be of a broader variety too: values, curves, areas, maps, images, alerts.

In essence, "game-like" interfaces provide an opportunity for the modeler to expand the usefulness of his model. Rather than being just an answers generator, the model becomes a *tool* for the decision making, an interactive and pleasant-to-use *tool*.

4 IMPLEMENTATION:THE GAME IS NOT OVER

4.1 An Engine

To implement anything even remotely resembling real-time interactive interface one may start with the existing simulation game engine. Gaming companies readily sell or license their engines to third parties. The older the game the less expensive the engine: engines that are 2-3 generations old are quite affordable. In order to attract more potential buyers gaming companies make game engines fairy flexible and easily modifiable.

Game engines, as a conglomerate of programmable objects, while not designed specifically for *risk* modeling, can be made suitable for it. What's important is that the engine provides *both*: links to the rich interface and an environment for a model's implementation. In essence, by changing a few formulas in the engine and by renaming a few labels in the interface, one may convert a computer game into an interactive risk model.

4.2 Modifications

"Game-like" interfaces place demands on the underlying engine to provide enough information with enough detail to be rendered for visual presentation. Every popular game in a "city/empire" genre already has a quite sophisticated simulation engine. Risk models designers, however, may need to improve it on in several crucial areas.

4.2.1 Simulate economy

Evidently, games designers usually don't bother with supplying their economic models with real-life data: even less so with the process of updating these data and maintaining economic parameters up-to-date. One shouldn't expect a game to have a sophisticated interest rates generator or an accurate implementation of the corporate tax code. However, games **do** have economic simulation modules; they just have to be modified and improved.

4.2.2 Simulate company

Games simulate building, structures or empires. They do not simulate insurance or reinsurance companies, property/casualty losses or facultative treaties. What is encouraging, though, is the fact that a city is much more complicated entity than insurance (or even reinsurance) company.

4.2.3 Simulate correlations

Another challenge for the model designer is the simulation of correlated random processes. Possible geographic components may only add complexity to the problem.

4.2.4 Simulate the rest

Given enough information a designer may try to simulate and incorporate into the model other random processes such as competition, taxes, geography, weather and catastrophes. It is useful to know that "game-like" interfaces support, visually and interactively, all of these features.

4.2.5 Monitoring

Even if model designers were able to perform all the necessary research, collect all the necessary data and effectively implement all the necessary algorithms for simulation of all internal and external processes surrounding the life of an insurance company, there would still be a lot of work to do. For the model to become a "game-like" decision making tool, one has to decide what statistics to calculate, what indicators to observe, which criteria to use for issuing alerts. "Game-like" interfaces technically may accommodate any number of them. However, to remain truly useful and approachable the interface has to expose only *key* indicators and to issue only *critical* alerts.

4.3 Simulation engines

Luckily, the majority of proposed modifications don't have to be designed from scratch: they are already satisfactorily implemented in the existing packages. The technology can be licensed or borrowed, but the fact remains: it is all doable.

4.3.1 Risk Explorer™

Risk Explorer™ by Ultimate Risk Solutions includes a comprehensive macroeconomic model, an innovative correlation module and many other useful features.

4.3.2 Public Access DFA Model

Public access DFA Model provides insight into insurance company simulations: both on the liability and asset sides.

4.3.3 Custom sims

One can imagine that large and successful reinsurance companies, major rating agencies and hedge funds in one way or another have developed working risk simulation solutions. Inevitably their knowledge will become public and tapping into this resource may prove invaluable for the "game-like" risk model designer.

5 CONCLUSION: AND THE WINNER IS…

Modern simulation games evolved into all-encompassing virtual worlds with rich and interactive interfaces. Game designers proved that modern computers are capable of simulating multiple aspects in the life of complex entities such as a city while keeping the game attractive and **approachable**. At the same time risk modelers faced the problem of explaining the model's findings to decision makers. Eventual integration of game design achievements into the risk models seems inevitable.

Evidently, the productivity rises when a tedious activity is camouflaged as a game, but this is just one of the benefits of such a fusion. Implementation of the "game-like" interfaces for the dynamic models brings so much more to the table: animation, geographical localization, monitoring (with alerts) and, most importantly, **interactivity**. It may drastically emphasize the roles actuaries are playing: model builders, parameter suppliers, algorithm implementers. Their role in a decision-making process would become transparent and self-evident. For the management Dynamic Modeling may stop being a black-box mystery but rather a desirable topic for discussion if not the major instrument for decision making. Decisions would be made on more solid ground. Companies may become more profitable. Shareholders would be rich, policyholders would be happy. Everybody wins!

6 Credits

The author is thankful to Will Wright, Sid Meier, Brian Reynolds and other game designers who helped to create such simulation magnum opuses as SimCity, Civilization, Caesar and similar edutainment masterpieces. This article would be of a much lesser quality without in-depth discussions about risk modeling with Alex Bushel, Vladimir Ladyzhets and Yakov Lantsman.

7 P. S.

Paradigm shifts, significant changes in our system of self-evident truths, don't occur overnight. Rather they happen steadily as carriers of one conceptual worldview are gradually replaced by the carriers of the new way of thinking. Someday we will see an influx of people who will not wonder what does visualization and animation have to do with the actuarial science and for whom Gaming (a.k.a. interactive visual experience) will not only look natural in the boardroom, but will actually be used as indispensable decision making tool. In the meantime, the author bit by bit, module by module, article by article will try to materialize his vision of what he believes is an Ideal Actuarial System.

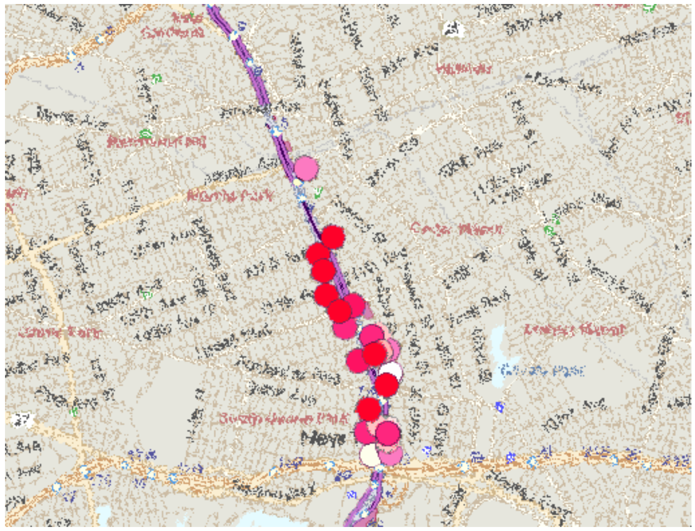
8 Links

[1] Daniel "Wolff" Dobson and Ken Forbus. *Towards Articulate Game Engines.* http://www.lggwg.com/wolff/aicg99/dobson-forbus.html
[2] Jake Simpson. *Game Engine Anatomy 101.* http://www.extremetech.com/article2/0,1558,1156338,00.asp
[3] Averill M. Law, Ph.D. *Eleven Critical Pitfalls in Simulation Modeling.* http://www.averill-law.com/simulation-news-papers.htm
[4] Stephen P. D'Arcy, FCAS, MAAA, Richard W. Gorvett, FCAS, MAAA, Thomas E. Hettinger, ACAS, MAAA, Robert J. Walling III *Using the Public Access DFA Model: A Case Study.* http://www.casact.org/pubs/forum/98sforum/98sf053.pdf
[5] Ultimate Risk Solutions. *Risk Explorer*™. http://www.ultirisk.com/home/aboutus.html
[6] Pinnacle Actuarial Resources Inc. *Public Access DFA Model.* http://www.pinnacleactuaries.com/pages/products/dynamo.asp
[7] Thomas S. Kuhn. *The structure of Scientific Revolutions.* http://www.emory.edu/education/mfp/Kuhn.html

9 Appendix: Expansion Pack

Visualization, animation and gaming are not such foreign notions for the insurance industry after all. Used suitably and cleverly these technologies may bring tangible benefits to areas as diverse as claims management, fraud protection and reserve testing. The author himself had an opportunity to use them in his own real life projects.

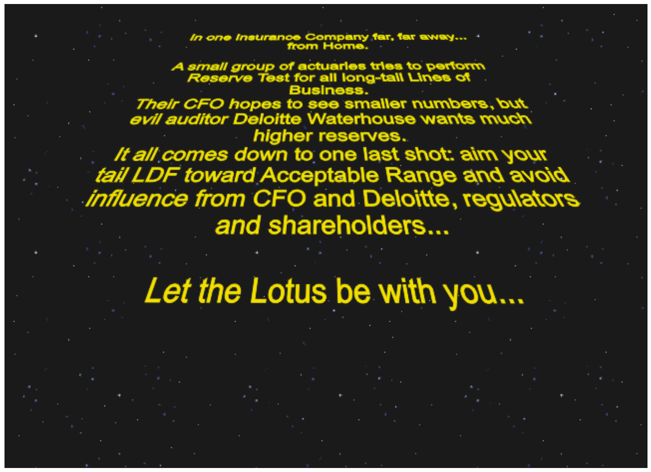
A. Studying GL losses from some treaty by looking at the triangles, vectors and raw data in databases didn't yield any obvious irregularities or suspicions. Everything looked normal until the data was **visualized** by placing the claimants' addresses on the map. The resulting picture (map labels are distorted on purpose) clearly showed an abnormal concentration along a highway.



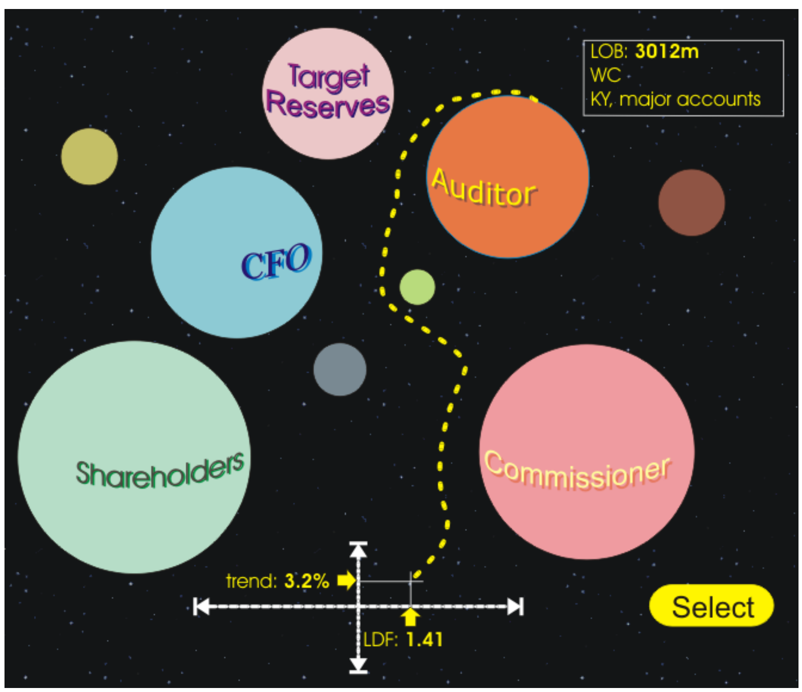
Apparently, these claims were from homeowners whose houses were damaged by the drillings on the nearby elevated railway construction. After combining all of these claims into one occurrence a company was able to recover a few million dollars.

B. Encouraged, this company ordered visualization studies on all of its treaties. Another interesting result was obtained with the help of **animation**. A treaty was found with all of its claims distributed quite homogeneously… but only in spatial dimensions. In a temporal dimension, though, there was a significant spike. During animation a huge chunk of claims appeared almost simultaneously (they suddenly "jumped" out all at once). After some investigation it was found that the majority of these claims were from car owners whose cars were damaged on the same parking lot by the cement pouring from the nearby construction site. Once again: one occurrence and several million in excess recoveries.

C. More than a decade ago the very first actuarial **game** (imaginatively named *"Tail Factor"*) was created.



The player, by selecting a few parameters like trends and LDFs, had to shoot a photon torpedo from his *"Data"* X-wing space-fighter towards an *"Affordable Answers"* Death-Star target. The torpedo had to avoid a gravitational pull from multiple planets labeled *"Auditors"*, *"Policyholders"*, *"Shareholders"*, *"Chief Actuary"* and *"CFO"* making an actuarial selection in the game as difficult and controversial as in real life.



It was a very primitive game, but the metaphor somehow clicked: users were paying attention to development factors and selecting medical cost inflations more thoroughly than they ever did. Selecting a tail factor in a "game" apparently was much more fun than doing the same on the dreary Lotus screen.

Biography of the Author

Aleksey Popelyukhin is a Vice-President of Information Systems with the 2 Wings Risk Services in Stamford, Connecticut and a Senior Vice-President of Technology with the Sam Sebe LLC. He holds a Ph.D. in Mathematics and Mathematical Physics from Moscow University (1989). Aleksey is presently developing an integrated pricing/reserving/DFA computer system for reinsurance and also an action/adventure computer game tentatively called "Actuarial Judgment." Dr. Popelyukhin is an active member of several scientific societies and an author of almost 20 scientific publications. His article "The Big Picture: Actuarial Process from the Data Management point of view" (1996) won the prize for the Data Management Discussion Call Paper Program in 1997.

Actuarial Data Management In A High-Volume Transactional Processing Environment

Joseph Strube and Bryant Russell, Ph.D., ACAS, MAAA

Abstract

The development and management of data resources that support property/casualty actuarial work are very challenging undertakings, especially in a high-volume transactional processing environment. In order to equip actuaries with the data resources necessary to excel in the performance of their functions, an Actuarial Data Management (ADM) support team is needed. It serves as a proactive, added-value conduit of business data and specialized technical support to an actuarial staff.

This paper examines the evolution of the actuarial data management function in the context of end user computing, and highlights the key roles and processes that comprise an effective data management operation in a modern property/casualty actuarial department. The paper also includes a case study that describes the development of the data management function in the Actuarial Department of Motors Insurance Corporation, a member of the GMAC Insurance Group, located in Southfield, Michigan.

Keywords. Actuarial Applications and Methodologies; Data Management and Information; Actuarial Data Management; Actuarial Technician; End User Computing; Data Warehousing; Insurance Data Management.

1. INTRODUCTION

The goal of an Actuarial Data Management (ADM) unit is to equip an actuarial staff with the data resources necessary to excel in the performance of their functions. Having this goal, it is clear that data managers and their crews fill a service role. But they are equally manufacturers too --- of information products. They construct focused relational databases from ever expanding data warehouses, and multi-gigabyte, multi-dimensional data arrays from voluminous mainframe transactional processing systems. And on a less grand scale, they even electronically whittle desktop data files after much sifting, selecting, and aggregating.

The types of processes that render data products are really not new over the past 30 years, but the tools used to retrieve and manipulate the growing stores of raw data surely have evolved. Actuarial analysts have always needed to extract the data from wherever it could be excavated. People with longevity in this field will no doubt recall a manual preparation process using tabular worksheets on 15-column ledger paper. More fortunate actuarial technicians would copy numbers from “DP” (Data Processing) Department green bar reports, transcribing them into cascading loss triangle ledger pages. Less fortunate ones would need to derive the incremental differences before penciling in the new month’s or quarter's development of counts and amounts. Once the rote work of initial spreadsheet updating was completed, more derived data was generated. The spreadsheets from the previous process provided the input for the next series of manually computed and manually recorded numbers. This time the computations resulted in values approaching analytical data: frequencies, severities, and pure premiums. Eventually, the assembly line of data worksheets and computations concluded in a hand-off to the actuaries who would analyze the refreshed statistics and interpret their meaning in reference to their pricing recommendations, and the company's loss reserves and premium adequacy levels.

In the context of today's actuarial activities, this chain of events continues. But the details of the process are so radically different that anyone familiar with a modern actuarial department could vehemently debate the notion. The skill set required to prepare the data for actuarial analyses has escalated from a quasi-clerical level to an advanced blend of technical and business acumen. Pointing out the vast difference in those skill sets, however, doesn't detract from the valuable data quality and auditing contributions made by the technicians of the era preceding end user computing. But it can serve to highlight the multi-faceted role that a modern data technician must fulfill as the focal point for coordinating the data extraction, cleansing, transformation, aggregation, and deployment processes within a company's actuarial operations.

And that brings us to the crux and purpose of this paper. The incredible volume of raw data that is available to end users through the business processes and information technologies employed in large insurance companies today, requires an orchestration of human and technological resources to achieve the goal of equipping actuaries with the data resources necessary for them to produce excellent results. This paper will discuss the roles of the human resources, the data handling processes, and the technologies that are key to achieving an effective actuarial data management function in a high-volume transactional processing environment. For the purposes of this discussion, a high-volume transactional processing environment (HVTPE) is one where transactions added to the most granular actuarial data resource in the organization exceed one million per month.

2. BACKGROUND

2.1 The Evolution Of End User Computing

By the early part of the 1980s, end user computing had emerged as a technology field all unto its own. The college graduates of the '70s who majored in business, mathematics, and both the physical and social sciences were trained in their curricula to incorporate computers into their research, data analysis, and presentation tasks. Earlier on, this was limited to time-sharing on large remote mainframe computers. And as the decade advanced, mini-computers appeared in American business and industry. These smaller scale machines made computing more accessible to onsite personnel because the manuals to operate and program the computers were generally available to the business-based technicians that were capable of using the equipment. Further, the input/output media used by these computers (particularly 8" diskettes and data cassette tapes) were readily available to the non-Information Systems class of people in the company. The day had arrived when business users could input, retrieve, and record their own work, in essence restoring a degree of local control that they had taken for granted in the totally paper-based office. This issue of control would prove to be one that would resurface as a point of contention between business professionals pushing for greater computing capacity and the technology professionals that managed the hardware and software computing resources.

In larger insurance organizations where monthly policy, premium, and claims activity required a company's computer(s) to handle millions of transactions, Information Systems (IS) personnel were continually compelled to become familiarized with the next generation of mainframe CPU. Their task was to gauge when the next upgrade of hardware and software would be needed to deal with the burgeoning flow of data that the business side of the company said was necessary to achieve its sales, marketing, financial, and management goals, as well as meet regulatory requirements. A lack of vigilance to the company's computing growth requirements or a misjudgment of a technical solution's scalability could result in a crisis of "computer resource gridlock." This gridlock would be manifest with the off-hours (evening) batch processing jobs contending for more time than they were allotted, threatening the daytime online processes that supported the field organization's customer service activities.

Add to this ongoing struggle for equilibrium, a business end user community that was growing impatient with the continuing dependence upon the IS area for obtaining large data compilations and department specific computer applications. Actuarial end users, with their more analytically inclined orientations, were at the forefront of the push to put increasing amounts of computing power, data storage capacity, and next generation (end user) programming tools into the hands of people outside of the IS Department. Non-technical business professionals became end user activists, seeking to control their own departmental systems development and data management destinies. It was an unsettling epoch of events for the caretakers of a corporation's computing resources. Before that time the domain of systems programming was entered into only through formal education in computer science at the collegiate level. Now "unqualified" non-technicians insisted that some of the company's IS budget be spent on computing resources that they themselves would utilize.

2.2 End User Roles Emerge

As end user computing became commonplace in the insurance industry, users began to coalesce into groups differentiated by varied levels of technical interest and ability as well as functional role. In a research study conducted by John F. Rockart and Lauren S. Flannery of the Sloan School of Management at MIT, 250 people involved in end user computing from seven organizations (three Fortune 50 manufacturing companies, two major insurance companies, and two sizable Canadian Companies), yielded six distinct types of end users.[1] The profiles of these six user types (Non-Programming End Users, Command Level Users, End User Programmers, Functional Support Personnel, End User Computing Support Personnel, and DP Programmers) are displayed in Table 1. These categories of end users were defined in 1983 when the practice of end user computing was growing at a rate of approximately 50-90% per year in the organizations included in the study.[2] And they were predicated on the then current mode of end user computing --- mainframe-based computer software tools and data storage.

TABLE 1: Rockart & Flannery's Six Types Of End Users

Non-Programming End Users	Only access to computer-stored data is through software provided by others. They neither program nor use report generators. Access to computerized data is through a limited, menu-driven environment or a strictly followed set of procedures.
Command Level Users	Have a need to access data on their own terms. They perform simple inquiries often with a few simple calculations such as summation, and generate unique reports for their own purposes. They understand the available database(s) and are able to specify, access, and manipulate information most often utilizing report generators and/or a limited set of commands from languages such as FOCUS, RAMIS II, EXPRESS, SQL, or SAS. Their approach to the computer is similar to that of an engineer to a slide rule in days past. They are willing to learn just enough about the database and the software to assist the performance of their day-to-day jobs in functions such as personnel, accounting, or market research.
End User Programmers	Utilize both command and procedural languages directly for their own personal information needs. They develop their own applications, some of which are used by other end users. This latter use is an incidental by-product of what is essentially analytic programming performed on a "personal basis" by quantitatively oriented actuaries, planners, financial analysts, and engineers.
Functional Support Personnel	Sophisticated programmers supporting other end users within their particular functional areas. Individuals who, by virtue of their prowess in end user languages, have become informal centers of systems design and programming expertise within their functional areas. They exist today as "small pockets of programmers" in each functional organization … Provide the majority of code for the users in their functions. In spite of the large percentage of time that these individuals spend coding (several estimated over 80%), they do not view themselves as programmers or data

	processing professionals. Rather, … [their] primary task is providing tools and processes to get at and analyze data.
End User Computing Support Personnel	Most often located in a central support organization such as an "Information Center." Their exact roles differ from company to company. Most, however, are reasonably fluent in end user languages and, in addition to aiding end users, also develop either application or "support" software.
DP Programmers	Similar to traditional COBOL shop programmers except that they program in end user departments wishing to hire "contract programmers," to avoid high consultant/programmer fees, and to build a larger base of knowledge of end user language computing within the corporation.

2.3 End User Technology Evolves

From those early years until the present, the evolution of the technology supporting end users is nothing less than astonishing. As the mainframe tools and data storage devices matured, the introduction of stand-alone personal computers and simple stand-alone application software presented a new dynamic. These personal-level computing devices supplemented end users' arsenal of capabilities for processing, analyzing, and presenting data. PC's were connected to mainframes to facilitate the transference of data from one platform to the other. Stand-alone personal computers were connected to each other as well as other "sharable" resources (like printers and hard disk drives) to form local area networks (LANs). PC/LAN hardware and software escalated in throughput capacity and functionality. Simple "peer-to-peer" networks were replaced by "client/server" networks in which a single more powerful computer served up software and data files to less powerful remote units. The "graphical user interface" style of software, using a desktop metaphor on the PC screen and point-and-click icons, became the routine expectation when powering up one's personal computer. LANs were connected to other LANs in geographically remote office locations to form Wide Area Networks (WANs). The continuing cycle of hardware capacity growth followed by software functionality growth swept across the business world in waves that incremented in amplitude and increased frequency year after year. And as this phenomenon protracted, entirely new branches of data management technology erupted from the trunk of traditional Database Management Systems (DBMS) technology, e.g., Data Warehousing, Data Marts, Online Analytical Processing, Extraction-Transformation-Load Software, Metadata Repositories, Decision Support Systems, Data Profiling/Cleansing/Integration Software, and Data Mining.

Each of these new technologies advanced particular aspects of the relatively static processes of data collection, processing, storage, retrieval, and analysis. A general description for each of these Data Management terms is provided below.

- Data Warehousing refers to the development and maintenance of a collection of large databases created from (and separate from) an organization's primary business processing systems. These large databases address enterprisewide subject areas and provide flexibility for management and business analyst reporting requirements. Being separate system constructs, data warehouses do not impact the operation and performance of the business systems supporting the marketing, sales, product support, and customer service activities of a company.

- Data Marts are smaller, specialized databases, often created from data warehouses. They focus on particular department-level needs or a subset of subject areas.

- Online Analytical Processing (or OLAP) refers to a type of software that provides very rapid access to data stored in a database and enables users to view and analyze data as multi-dimensional arrays.

- Extraction-Transformation-Load Software refers to a highly specialized and powerful class of software that can perform data extractions from multiple source databases, translate and convert the data according to business rules specified by business and data analysts, then load the data to a target database structure for subsequent querying and analytical reporting purposes.

- A Metadata Repository is a special type of database containing information about another database, e.g., how the data in the other database was collected, transformed, and formatted, how frequently it is updated, and generally anything that can be useful to analysts that need to query data from that database.

- A Decision Support System (DSS) is a specialized database and an associated set of software tools that are dedicated to enabling management decision making processes.

- Data Profiling/Cleansing/Integration Software is a powerful class of software used to examine a number of characteristics of data in source databases, apply customized business rules to maintain or enhance the data's integrity and usability, and consolidate multiple occurrences into a "best" version for loading into a single target database.

- Data Mining refers to a highly sophisticated class of database applications that detect hidden patterns and relationships in a collection of data in order to predict future behavior.

2.4 End User Roles Adapt

The extraordinary technological advancements over the past 20+ years have required a commensurate growth in the knowledgeability and skill sets of professionals participating in the discipline of actuarial data management. And depending upon the role or roles a person has needed to fill in the spectrum of end user types, the limits of his or her mental elasticity have been unequivocally tested. Given the incredible growth of transactional data to mammoth proportions (hundreds of gigabytes, even terabytes in size), and the diversification and elongation of data management processes (capture, inspect, cleanse, transform, qualify [create metadata], integrate, store, and distribute), there are substantially more steps and more time required to manufacture and deliver data ready for an actuary's consumption. It will be helpful at this point to update the characteristics of the end user types (described in Table 1) for current practices. It is a subset of these user types, operating as an efficient conduit of technology, which comprise the key roles necessary for truly effective actuarial data management. Refer to Table 2.

TABLE 2: Updated Characteristics of End User Types

Non-Programming End Users	Only access to data is through data products delivered by others. They do not program. Access to data is through menu-driven computer interfaces, standard desktop tools, or a strictly followed set of procedures.
Command Level Users	Have a need to access data on their own terms. They perform simple inquiries often with a few simple calculations such as summation, and generate unique reports for their own purposes. They understand the available database(s) and are able to specify, access, and manipulate information most often utilizing report generators and/or a limited set of commands from database query tools.

End User Programmers	Utilize command and procedural languages, as well as database query and creation tools directly for their own personal information needs. They develop their own applications and databases, some of which are used by other end users. This latter use is an incidental by-product of what is essentially analytic programming performed on a "personal basis" by quantitatively oriented actuaries, planners, financial analysts, and engineers.
Functional Support Personnel	Sophisticated programmers and data handlers supporting other end users within their particular functional areas. Individuals who, by virtue of their prowess in end user languages combined with an expert knowledge of the business data, have become informal centers of systems design and programming expertise within their functional areas. Provide the majority of data for the users in their functions. In spite of the large percentage of time that these individuals spend coding and managing data for others, they do not view themselves as programmers or data processing professionals. Rather, their primary task is utilizing tools and processes to get at, prepare, and distribute analytical data to their functional area.
End User Computing Support Personnel	Most often located in a central support organization such as a "Help Desk." Their exact roles differ from company to company. Most, however, are reasonably fluent in end user languages and assist end users with troubleshooting data retrieval and reporting issues.
DP Programmers	Similar to traditional COBOL (procedural) shop programmers except that they program in end user departments wishing to hire "contract programmers," to avoid high consultant/programmer fees, and to provide technical advice and support for data retrieval, transformation, and distribution. Usually specializes in one or more data handling tools, e.g., Database Systems, Data Warehousing, Extraction-Transformation-Loading software, OLAP, and Data Mining.

Generally, the emphasis of each role has shifted from being a consumer or producer of a computer application, to being a consumer or producer of a data product. This may seem like an unnecessary nuance to cite considering the usual byproduct of a computer application is a data compilation or hard copy report of some sort. But the advancements in client/server and desktop PC applications have made the task of accessing and working with data relatively elementary, even for novice business analysts. Consequently, data files of varied sizes and formats are now a norm as deliverables to actuarial analysts.

2.5 The Alignment Of End User Roles And Business Roles

In a large insurance carrier, one or more particular groups of people can be associated with each of the end user categories described in the updated table. This is seen as a natural development based upon a combination of business function, professional preference, and technical/analytical ability.

The Non-Programming End Users are typically those in management roles, executives and senior managers, who have others prepare data and presentations for their review. These end users may also include corporate educators and process modelers who use the data as presentation or communication devices.

Command Level Users include business analysts who have learned how to access the organization's data warehouse and, if available, selected data marts. Their use of data retrieval and reporting tools continues to be limited to what's essential to conduct their day-to-day jobs in areas such as human resources, accounting, market research, claims, and field operations.

End User Programmers continue to include people who are quantitatively oriented and utilize computer and data resources to build reliable and repeatable processes for satisfying their own information requirements in the context of their job responsibilities. These people are actuaries, actuarial analysts, planners, and financial analysts.

Functional Support Personnel include people who are individually a blend of businessperson and technician. Their role dictates they provide support for their department of residence and may be comprised of people who previously filled the roles of actuarial or financial analysts, but due to a technical expertise that developed over time, they transitioned into a specialized, quasi-technical position. In addition, this group can include entry-level personnel that are training in the data rudiments of the analytical and professional roles they are aspiring to fill. These people include actuarial technicians, financial technicians, data analysts, data technicians, as well as interns or college co-op students.

End User Computing Support Personnel include technical analysts that interface with the end user community to answer questions, act as points of contact for troubleshooting issues, provide or augment training, and manage user access to the organization's corporate level data resources (the data warehouse and cross-functional data marts). This group may be known as some type of Help Desk or User Hotline, or associated with the technology employed by end users, e.g., the Data Warehouse Support Group or the Decision Support System Team.

Lastly, the DP Programmers include internal or outside contractors that supplement and extend the technical competencies of the other end user categories on an as-needed basis. These people are often brought in to fill roles that require specializations outside the normal technical parameters of on-staff business end users.

3. KEY ROLES IN THE MANAGEMENT OF ACTUARIAL DATA

3.1 The Critical Trichotomy

All the end user categories described above are evident in either the administration of or the execution of actuarial data management processes at an insurance concern with a high-volume transactional processing environment (HVTPE). And more than one end user type can apply to different actuarial and actuarial support functions. There may be some crossover of categories between positions as well. For example, actuarial management typically cluster into the Non-DP Programmer and Command Level User groups, actuaries and actuarial analysts to the Command Level and End User Programmer groups, actuarial technicians to the End User Programmer and Functional Support Personnel groups, and lastly, help desk staff to the End User Support and DP Programmer groups. Technical contractors may function in a number of capacities that mimic the roles of functional support staff and help desk personnel besides filling specialized consultant roles.

To varying degrees, these functions each contribute to the development, maintenance, and management of effective and excellent actuarial data resources. Yet two of these functions, in conjunction with support from the company's Information Technology (IT) Department, are more crucial to the ongoing actuarial data environment than the others. In the modern insurance enterprise, it is the actuary, the actuarial technician, and the IT management function that form a critical foundation for an effective actuarial data management operation. Each of these three functions brings unique and indispensable elements into an alliance dependent upon mutual cooperation. This trichotomy will now be more fully described so as to compose an interaction model for accomplishing corporate actuarial objectives.

3.2 The Role Of The Actuary

3.2.1 HVTPE – The Good News And The Bad News

A high-volume transactional processing environment (HVTPE) presents the actuary with a classic good news/bad news story. The good news is that the actuary can rely on an extensive historical view of past transactions. The bad news is that the past transactional experience is extensive!

3.2.1.1 The Good News: A Lot of Good Information Available

Having a large volume of potentially credible data enhances the actuary's ability to analyze a company's historical experience.

- ♦ **Company-specific vs. industry experience.** The relationships between different classes, territories, etc. of the company's books of business can be examined by summarizing the data along various common dimensions. Company experience is readily comparable to industry-wide summary data.

- ♦ **GLMs.** The HVTPE is a Generalized Linear Model user's paradise. The information is available at a very granular level, and with proper summary, the GLM can be employed to compare multiple factors that bear on the performance of the overall book. Perhaps territory and deductibles are interactive. The GLM allows the actuary to consider both dimensions (and others) in one analysis.

- **Loss Distributions.** Loss distribution analysis is another area that relies on a highly granular data resource for detailed historical size of loss experience. The HVTPE is a particularly attractive resource for this type of statistical analysis. The sheer volume of historical data may permit the actuary to meet the parallel demands of homogeneity and credibility by restricting the view to aggregations of subsets with similar loss potential.

- **Reconciliation between functional areas.** Because the HVTPE provides a very granular level of detail, it becomes possible to directly compare actuarial aggregations (e.g., accident year, policy year) to other financial aggregations (e.g., calendar year, calendar quarter, underwriting year). This comparison can lead to a closer working relationship with other (non-actuarial) areas of the company. The various areas of responsibility no longer need to focus on which data is "right," but rather on what the various aggregations of data suggest about the book of business.

- **Cross-functional use of common data.** The actuary's requirements for data from a HVTPE will tend to be fairly detailed. In many cases, other non-actuarial areas of the company (e.g., claims, underwriting, and marketing personnel) find the resulting body of experience a valuable resource for their needs as well.

3.2.1.2 The Bad News: The Actuary May Spend Inordinate Time On Data Management

The HVTPE requires an "industrial strength" toolbox of hardware and software tools to store, aggregate, retrieve, and analyze the data. Consequently, "someone" needs to be concerned with the following items:

- **Hardware and software platform specifications**
- **Hardware and software procurement, maintenance, and updates**
- **Balancing**
- **Monitoring and maintenance of information quality and consistency**
- **Production vs. ad hoc environment**

Certainly, some actuaries are capable of providing the guidance for managing these elements of the HVTPE process, yet this is not the best use of the actuary's skill set. This data management role is better handled by a professional partner, the actuarial technician, working in tandem with the actuaries.

3.2.1.3 Involvement In The Development Of The Data Resources

It might be tempting to the actuary to place the responsibility for designing and implementing the actuarial data management processes of the HVTPE completely in others' hands. However, this is not realistic for many reasons:

- **The actuary cannot rely purely on canned reports.** Predetermined summaries of the various books of business, subsets, rating variables, etc., are helpful. However, every question raised by such reports will suggest further "deep dives" into the historical experience. If the data retrieval design is not clearly set out beforehand, this may result in an excess amount of time being spent on creating ways to get at the additional data. In the worst-case scenarios, the information being sought may not be accessible to the actuary.
- **Information that provides actuarial value may be seen as of secondary importance to other business areas.** For example, historical policy rating information may not be seen as critical to the sales area, and current policy rating information may not be of much use to the claims personnel administering claim payments. The pricing and reserving actuaries, on the other hand, would find value in both historical and current policy information. Actuarial involvement helps ensure that both types of information are made available for actuarial use.
- **The required level of detail for actuarial analysis is different.** In many cases, the actuary will require information that is more summarized than individual policy or claim level, yet is far more detailed than the financial reports required for management review of the business. Relying exclusively on operation reports (too detailed for actuarial analysis) or management reports (too summarized for actuarial analysis) prevents a full actuarial review of the programs.
- **Historical retention of useful data is important.** The actuary must help determine how long information should be made available for actuarial analyses. There is a balance between establishing the long-term patterns in claims emergence versus retaining data that no longer has any reasonable relationship to today's book. It is not just active policies and claims, nor just open tax years. The actuarial information requirements would most likely include these time frames and more. The actuary is best suited for determining this balance.
- **"User defined" information fields are needed.** Some actuarial aggregations involve both historical information and "user defined" information not available in other areas. For example, the definitions of "territory" and "symbol group" can change over time. Hence, the historical information may need to be re-rated or re-cast with the current or prospective view of the data. Such views are important for

actuarial analysis, yet of lesser importance to the other areas of the company. Without actuarial involvement in the definition of requirements, there is a risk of losing this capability.

- **Changes to the HVTPE have consequential impacts.** Actuarial informational requirements can be seen as a "next phase" item, rather than an upfront requirement. It is a natural tendency to ensure that all of the operational data needs are met before discussion of "back-end" reporting begins. However, many of the initial requirements set the precedence for long term informational deliverables. The order of delivery priority, i.e., delivering the operational needs first, is correct. However, gathering all of the informational needs, including the actuarial informational needs, should be done up front. Otherwise, there is a great risk of losing the opportunity to capture and incorporate extremely important data elements into actuarial analyses. Therefore, it is important for the actuary to be involved in this upfront process. This means a partnership is needed between the actuarial and non-actuarial areas that initiate projects that alter the HVTPE.

3.2.2 Qualifying Actuarial Data Requirements

The actuary relies on the ADM team to implement and maintain a successful system for capturing, storing, and retrieving HVTPE information. The actuary must clearly define and communicate to the ADM team specific data requirements.

3.2.2.1 Historical Data Requirements

The HVTPE information source can be extremely large. The actuary needs to start by identifying those components of the historical information that need to be captured and available. Typical items include those policy characteristics that shape the premium rate, as well as the claim characteristics that have a bearing on the claim size, frequency, emergence, etc.

3.2.2.2 Level Of Detail Requirements

Beyond "What data?" is the question of "How detailed?" This can be a very intensive, time-consuming effort. Does it make sense to go below "claim" level? What is a "claim"? If several transactions comprise a "claim," how important is it for the actuary to be able to combine the cumulative flow of transactions at the claim level? The same questions apply to multiple transactions that comprise events at the policy level. Together, the actuary and ADM team must consider the trade-offs between maintaining detail at a more detailed level versus the effort necessary to capture, store, maintain, retrieve, and aggregate the information.

3.2.2.3 Required Time Frames

Actuarial time frames of interest include transaction dates (e.g., report date, process date, accounting dates, etc.) as well as more intrinsic dates (e.g., loss incurred date, underwriting year, calendar year, etc.). Whenever possible, transaction dates should be stored explicitly. The reason should be clear, as information that "everyone knows" today becomes "no one remembers" as time passes. Storing the date information explicitly ensures that historical data retains its meaningfulness and its place in time as the data ages.

3.2.2.4 Aggregations And Summary Level Requirements

Some HVTPE information provides value at a very granular level, e.g., location of claim, while other information must be summarized to be of value. Examples include loss experience by class, premium by territory, etc. The ADM team can work to ensure that the data resource aggregations are achieved in an efficient manner.

3.2.2.5 User-Defined Fields That Change Infrequently

Certain user-defined information can be computed and stored along with the historical values themselves. Examples might include geographical regions built from ZIP codes, descriptions of deductible codes, and aggregations of business accounts into broad categories. The benefit is two-fold: a consistent definition and efficient summarization of key business information. The trade-off is the upfront time and effort required to build and store this information. This type of effort is best suited for information that does not change frequently. For example, building "state" from ZIP code is a fairly static, well-defined computation. The actuary can improve the usefulness of the data by requiring this information be pre-computed and available for retrieval.

3.2.2.6 User-Defined Fields That Change "On-The-Fly"

The actuary should require the ability and capacity for building summaries based on data fields that can be changed "on the fly." This is particularly necessary when the classes of business are periodically reviewed and re-classified. If the user-defined aggregations change, the assignment to each data element extracted from the HVTPE may change. For example, newly introduced vehicles are typically assigned to a vehicle class based on judgment. Once loss information has become available on a vehicle class, it is not unusual to find certain vehicles re-classified. The actuary may then analyze experience by <u>currently assigned</u> vehicle class, not by historical class.

The actuary's data is not purely historical in this instance. It is a combination of the historical experience extracted from the HVTPE (e.g., loss experience), combined with a user-defined aggregation (e.g., current vehicle class) that is derived from the historical data.

3.2.2.7 Support The Value Proposition Of The Data Dictionary

A data dictionary ensures a consistent view of what the information means across the organization. It provides a precise definition of what the field is called, what information is expected to be stored in the field, what the typical values for such a field might be, etc. The Data Dictionary can also document when a data field has only recently become available with useful information, or whether another data field has ceased to be populated with current information (e.g., if a program has been placed in run-off).

3.2.2.8 Historical Data Retention

The actuary's retention period for historical data will most likely be different from those required for other parts of the organization. Actuarial analysis may require use of policy year, calendar year, accident year, etc. For longer-tailed business, the aggregations may occur over many years of data. The actuary must balance between storing too little history, and storing more data than necessary, impairing the ability to efficiently retrieve useful information.

3.2.3 Critical, Actuarially Valuable, And Nonessential Data

The actuary must distinguish between information that is critical to the actuarial analysis, versus information that has potential actuarial use, versus nonessential information.

3.2.3.1 Critical Data Elements

Without the critical data elements, there is no reason to pursue construction of a distinct actuarial data solution derived from the HVTPE. Examples include loss information in sufficient detail, critical dates (e.g., loss incurred date), premium amounts, etc. An information system lacking these elements cannot provide sufficient detail for a full actuarial analysis of the corresponding business programs.

The actuarial data requirements for critical data elements must be clearly communicated to the responsible parties. It can be a fatal flaw in a HVTPE project to just assume that "everyone" is aware of the actuarial importance of this type of information.

3.2.3.2 Actuarially Valuable Elements

Other data elements can serve as valuable input to future actuarial analyses, yet cannot be considered critical. These data elements can be thought of as "actuarially valuable." Without these data elements, the actuarial landscape is bleaker and the analyses are thinner. Yet, the remaining actuarial information set will allow actuarial analysis to continue in some lesser capacity.

Because "actuarially valuable" lies one step removed from "critical," there will always be a question of whether or not it is justifiable to capture these additional fields. There is a trade-off between the cost of gathering, storing, and reporting on actuarially valuable fields, and the potential "what if" insight the data can provide. The actuary must be prepared to discuss what might be the potential value of each additional field captured.

The ultimate value of considering these data elements is in allowing the actuary to be proactive rather than reactive. The actuarially valuable data elements may not be examined in every analysis, but there is a time-to-market advantage in incorporating readily available information when needed. The additional data is helpful when supporting the introduction of revised rating systems. The actuarial valuable data elements also tend to show additional value when the actuary needs to investigate how or why a book of business deviates from its projected values.

3.2.3.3 Nonessential Data Elements

Finally, there are nonessential data elements. These are perhaps "nice to have," but not cost-effective from an actuarial point of view. The data elements are not crucial to actuarial analysis, or the potential value in an actuarial analysis is limited. For example, the color of a vehicle could be examined as to its interaction with claim frequency or severity, but it is hard to imagine how this might be incorporated into an actuarial pricing, rating, or reserving analysis.

It might still make sense to consider nonessential data elements, if the information is of value to other areas of the company. For example, claims adjuster ID, or sales agent number, may be valuable to another area. Inclusion in the HVTPE data extract would allow other areas to leverage the actuarial effort.

3.3 The Role of the Actuarial Technician

3.3.1 The Dual Roles

As a pivotal member of an Actuarial Data Management unit, the Actuarial Technician must always be focused on their top priority: ensuring client actuaries are provided with the data resources necessary to excel in the performance of their functions. This involves two types of support roles: one as a data facilitator, the second as a data supplier.

3.3.2 The Data Facilitator Role

As a data facilitator, the Actuarial Technician regularly monitors the corporate data resources that the actuaries depend upon, whether directly or indirectly. As issues arise related to the availability, accessibility, and integrity of the data, they are then in a position to advise the actuaries accordingly. Such advisories may be limited to simply notifying affected individuals about problems or pending circumstances. For example, these notifications would be made when data anomalies are observed in the data warehouse, data marts, or their source systems, or when the release of new data is accelerated or delayed. In other cases, the advisories may involve the relay of specific actions that must be taken to work around problems that have yet to be permanently resolved. Examples of these would include notifications to select certain instances of data elements over others due to problems that had surfaced, or perhaps providing details for filtering the data differently to avoid erroneous results. The point of these facilitating actions is to promote effective methods of obtaining the highest quality data possible, as well as enhance the productivity of the actuaries. Technicians can obviate the need for repeating tasks (or even whole processes) performed by the actuaries simply by being attentive to corporate data issues and circumstances surrounding the data systems and then conveying related information in a timely manner.

As a data facilitator, the Actuarial Technician also serves as an intermediary agent between the actuaries and the IT Department. Because the Technician's job is so dedicated to providing data to the actuarial staff, he or she is in a unique position for tracking the evolving needs at the local level. They are witnesses to the ongoing development of information requirements by virtue of being the only "first-tier supplier" of actuarial analytical data that is in close proximity to the actuarial consumer. This familiarity with data requirements is of particular value when new data requirements need to be conveyed to the IT area and formalized into a project. As discussed under The Role of the Actuary, data requirements must be determined by the actuaries; however, it is the task of the Actuarial Technician to facilitate the transfer of those requirements to those that must satisfy them when the supplier is the in-house IT Department or an outside technology vendor. The requirements transfer may involve editing documents, drafted by the actuarial staff, to expound on, clarify, or provide examples of the data being requested. Or it may involve providing interpretation of the request in terms that address corporate-mandated procedural requirements, e.g., forms and support documentation. If the data project deliverables are of a size or complexity requiring a phased in approach, then the Actuarial Technician should consult with the requestor(s) and discuss what options are offered by the supplier. And if any particular advantages or drawbacks among the options are evident to the technician, he or she should make the requestor(s) aware of the observations. Likewise, additional alternatives should be discussed if they would better serve the need and would be plausible for the supplier to accommodate. When the preferred options and priorities are decided, the Actuarial Technician should convey them to the supplier or facilitate a meeting of all parties. In the communications between the Actuarial area and the IT function, the Actuarial Technician is not impartial. He or she primarily represents the interests of the actuaries. With that said, when it becomes apparent that progress can only be made through a compromise of all parties' concerns, the Actuarial Technician should do their best to mediate a solution that achieves a balance between the contending positions without undue compromise to the actuarial position.

The last aspect of the Actuarial Technician as a data facilitator involves software and hardware tools. Again, because the technician's position is semi-business and semi-technical, they are in an advantageous position for researching and evaluating data manipulation tools that would achieve greater productivity not only for the ADM unit, but the actuaries as well. Database management tools were once considered to be exclusively within the IT staff's domain. But as the need for greater data manipulation capability evolved, these tools were adopted by the more technically-inclined end users, and eventually mainstream end users. This progression implies that the Actuarial Technician will remain at the forefront of the dissemination of software tools that will enhance the capabilities and productivity of the actuarial staff at large. Consequently, the actuarial technicians in the organization should keep other actuarial staff members apprised of tools that may provide such benefits.

3.3.3 The Data Supplier Role

As a data supplier, the Actuarial Technician is the fulcrum that allows an actuary to leverage his or her analytical abilities. Without the Technician to intermediate between the state of the data as it is stored in the corporate systems and its transformed state needed for statistical analyses, the time and skills of actuarial professionals would be heavily taxed. In an organization with high transactional volume, the proportion of time used to prepare data for analysis versus the time used to perform an analysis easily shifts from an 80% versus 20% proposition, to one of 20% versus 80%. And the shift cannot necessarily be discounted as the transition from older legacy systems to more accessible data warehouses or data marts occurs. For as the raw data is wrestled from older systems and kneaded into more refined and accessible chunks, the progressive requirements for analytical aggregations, as well as successive drill-down capabilities, emerge. In fact, the refinements involved with these types of data may expand the number of processes supported by the ADM unit, rather than simply replace them, because of the data's increased scope and the actuary's heightened need for its ongoing availability and accessibility.

As data suppliers, Actuarial Technicians perform both end user production activities and end user development activities. From the initial release of data that fulfills the input requirements of an actuary's periodic analysis, an implicit expectation materializes that the same data will be provided in an updated form in the future. The expectation may be communicated early on as part of the original request, or it may take the form of a "one-off project" that over time seems to recur in a variety of incarnations. In any case, the Actuarial Data Management group needs to maintain a production schedule for developing and distributing the data it is routinely expected to provide. That is job one. Unless the actuaries can count on the consistent and timely delivery of the input data for their recurring analyses, any new development work on the part of actuarial technicians is meaningless. The data as a whole will lose credibility not as a result of any inherent inferiority, but due to the unreliability of its providers. This consistency of timeliness is as important as the completeness and accuracy of the data itself in order to achieve superior data quality. So the support of ongoing production work must be the prime directive for actuarial technicians.

And the support includes addressing issues that would threaten the fulfillment of that directive.

Beyond the production activities, actuarial technicians need to allocate time towards new development efforts as well. These take the form of in-department projects, as well as corporate projects. The in-department projects represent enhancements to the existing data resources that the ADM team manages as well as new data development. Corporate projects can directly impact the data resources maintained by the ADM group either by altering the data that is fed to them or by altering the hardware and software infrastructure that supports them. In either case, the Technician needs to be involved and attentive to any negative effects by specifying, if not also performing, adequate user acceptance testing (UAT). Maintaining an ongoing presence during the course of the project by attending status meetings can often alert the Technician to hazards and issues that could result in detrimental consequences to their data systems that would not be apparent from a review of the project's business and technical requirements documentation.

3.4 The Role of the IT Management Function

3.4.1 The Information Technology Perspective

The extraordinary growth and advancement of the technology industry has compelled IT departments across the business landscape to expand, reorganize and reinvent themselves repeatedly in an attempt to meet the requirements of their business unit customers. The scope of the IT Management function now spans nearly every part of the modern insurance organization. The emergence of new technology, the drive for incremental improvement in business processes, and competitive pressures have propelled this expansion. The aspects of the role needed for an effective actuarial data management operation, however, are not quite so diverse. There are particular IT responsibilities that provide the key elements of support for achieving actuarial data management objectives.

3.4.2 Managing The Existing End User Infrastructure

First and foremost, the IT function must ensure the availability and functionality of the existing business computing infrastructure. This means more than simply troubleshooting problems after they've been reported to a help desk by users, but rather proactively managing the infrastructure. Is local area network (LAN) monitoring software in place to detect and alert IT management personnel about network traffic spikes and extended high loads that can be traced to substantial data transmissions? If a wide area network (WAN) is part of the data management infrastructure, does it have a traffic monitor with notification triggers? When network disk storage reaches a 75-80% utilization threshold, are warnings issued to archive and free up space so as to avoid abrupt interruptions? Also, if the opportunity arises for business personnel to be involved with infrastructure plans, it can be very beneficial for providing input. Sometimes decisions are made by IT management to reallocate resources that appear to be "on average" under-utilized. Summary level monitoring reports that are commonly used for such decision-making don't always present a valid picture. The resources in question may actually be utilized heavily for short periods at weekly, monthly, or quarterly intervals. A reduction of throughput capacity across a network or on a data server could seriously constrain the efforts to prepare and distribute updated data resources on schedule. The voice of a business user in the forum of an infrastructure planning meeting can make the IT area aware of that situation and avert a potential crisis.

3.4.3 Infrastructure Renovations And Innovations

Secondly, the IT function should facilitate the advancement of the infrastructure in such a way that the actuarial data management function as well as the actuaries can take advantage of already-installed technologies in new ways or adopt newer technologies that increase functionality and productivity. This can occur by providing access to current software remotely through dial-in and broadband channels. Because data management processes can require several hours, it would be helpful to have remote access so subsequent processes can be launched after normal business hours if automatic triggers are not available.

Also, the IT function can be especially helpful if they maintain a program for routinely upgrading versions of both server and desktop application software. In regards to new types of user software, it may be unrealistic to expect the IT area to keep current on products that would be especially beneficial to actuaries, unless a specific problem or a functional deficiency has been communicated. However, the actuarial area may become aware of new or enhanced products that promise to add substantial value to either the analytical or data management processes of the department. At these times, the IT area should be invited to jointly investigate the potential. The Actuarial area can assess the <u>value</u> of the products in terms of their business requirements, and the IT area can assess the <u>cost</u> of the products in terms of their installation and technical support requirements. Together, both areas can determine if the costs are justified by the anticipated values.

3.4.4 IT Project Management

Lastly, the IT function is needed to assist with the execution of projects that either 1) Offload work to an automated system developed and maintained by IT professionals, or 2) Exceed the ADM unit's domain of control.

In the first case, some of the data resources and programs created by the ADM unit for the actuaries may over time become basically static structures. That is, the architecture and computations contained in them do not require updating. It becomes enough to simply refresh the data for new increments of time. Yet the refreshment process may require several intermediate processes that require many days or even weeks to accomplish. Even though the programming is sound, offloading the routine work to a new production system built by the IT area would allow the ADM group to place greater focus on the more volatile and actuarially esoteric requirements of the department. In fact, ADM-produced prototypes of data marts and programs that satisfy routine requirements would both prove the concept of an IT project as well as serve to meet the actuaries' needs on an interim basis. Such prototypes can also serve as the basis for the requirements of a formal system development project.

In the second case, where a project exceeds the ADM unit's domain of control (or the group's scope), the IT area must be engaged to enlist and manage the necessary in-house and/or outside vendor resources. A common example of such a project is the addition of new data elements to the organization's data warehouse or interdepartmental data marts.

4. PROCESSES

4.1 Commitment To Succeed

Planning, designing, building, and maintaining actuarial data resources that house millions of policies and claims is not a simple undertaking. This deserves to be stated explicitly even though the majority of readers who have persevered reading to this point would never assume otherwise. However, there is no shortage of consultancies that sincerely profess they possess the knowledge, skills, tools, experience, and human resources to promptly craft a silk database from a sow-system's ear. And that is not to say that truly qualified consultants couldn't perform such a metaphorical miracle for a needy and adequately-budgeted actuarial manager. But the miracle simply can't happen without the dedicated, time-consuming participation of the actuaries and their support technicians who must articulate the data requirements in definitive terms, facilitate the collection and communication of critical technical information to the consultants, be willing to discuss data issues at length, and judge which, if any, and to what degree, compromises regarding the deliverables can be tolerated. As the proverb goes, you will only get out of it what you put into it. Whether the actuarial area's data resources are products of strictly in-house efforts or of consultants hired from outside the organization, key members of the actuarial staff need to be involved to whatever degree it takes to bring the initial databases online as well as sufficient commitment to oversee their maintenance ongoing.

4.2 Formal Vs. Informal Approach

With that said, data resource development efforts involve a number of processes that are generic. One or more of these processes can be approached within a formal system design methodology and using specially designed software tools, or they can be approached informally through a logical and incremental approach. Regardless of the magnitude of the project, when the development team needs to serve simultaneously as an operations team and a production support team, the logical, incremental approach may be the only practical means to accomplish a project's objectives. This is because production work must be regarded as paramount. As discussed under the Data Supplier Role of the Actuarial Technician, development work must be subordinate to production work or the credibility of current and future data deployments is diminished. Having sufficient resources to segregate ADM people between exclusively production and exclusively development work teams is without question a luxury afforded by few (if any) P&C insurance organizations. Consequently, the flexibility of following a simply logical and incremental approach to implementing data projects by an ADM team should be regarded as the norm, rather than the exception.

An example of approaching a data project using a simply logical and incremental approach would be rebuilding or refreshing an existing database or data mart using a new process or set of tools. When the opportunity arises to introduce new software tools or enhance a data development and deployment process, the ADM team can integrate the new tool(s) or reprogram the existing process(es) as time allows and run it in parallel with the current processes. This ensures that new methods and processes meet the existing standards as a minimum requirement, and provides continuity with past practices and data quality levels. A successful parallel test as well as continuity with past data deployments are very important for gaining the acceptance of new data products by actuarial data consumers.

In the case where IT Management leads a project and manages the development resources that are independent of the ADM team, the formal project methodology approach is expected to be the norm. In that instance, the technicians that design and build the project's deliverables can work as dedicated resources focused on the new development activities without engaging in the risky practice of placing production priorities in contention with development priorities.

4.3 The Processes Of Data Management

At a deeper level than the approach and management of an actuarial data project, the processes involved with fulfilling the requirements will normally imitate, if not actually parallel, those of data warehousing processes. Consequently, a survey of these processes are presented below to familiarize actuarial personnel with them.

4.3.1 Data Modeling, Metadata, And The Data Dictionary

The beginning process of actuarial data management is the identification, qualification, and modeling of the data required for actuarial analysis. Until the data requirements are sufficiently identified, defined, and structure added to them (as is done with data modeling), the deliverables conceived in the mind of an actuarial requestor may be very different than that of their support technician. The identification and definition of requirements are initially expressed in the data terms, descriptions, and valid values associated with the data elements that the actuaries need. Ultimately, the data requirements will be expressed in terms of data models, metadata, and a data dictionary.

Data modeling is the identification, analysis and organization of data elements into logical and physical database designs. Some data modeling software simply provide a means to build logical relationship diagrams among data entities for documentation purposes. Microsoft's *Visio* application is an example. Other data modeling software goes beyond mere designing of a database to the creating of a complete physical data model once the logical design is finalized. *Erwin* by Computer Associates is an example of that type of software.

Metadata is data about data in a system or data structure. Each data element has particular attributes that uniquely characterize it, such as its name, definition, description, data type, data length, format, valid values, domain ranges for values, source files and source data elements. When data is processed through retrieval, cleansing, conversion, and transformation stages, there is technical process metadata that is applicable to data elements as well, such as the date processed and business rules applied to derive the stored value. Supplemental business metadata is also valuable to capture for knowledge workers. Examples include descriptions of how the data values have changed over time or descriptions of how the data enters the system initially (via automated means or manual entry), and descriptions of how upstream business practices may have impacted the values observed in the field.

A data dictionary is a tool for displaying metadata to business and technical personnel. A data dictionary is important for expediting the transfer of knowledge regarding the meaning of data values stored in the data fields. Without a consistent point of reference for describing the meaning of codes as well as the sources and derivations of data elements, any analyst will be hampered in their efforts to build accurate queries and effectively analyze the queries' results. Likewise, data management technicians need to have a firm grasp of the technical metadata in order to build complex extraction and transformation processes.

4.3.2 Data Extraction, Data Profiling, And Data Quality

Data extraction is the process of selecting and copying discrete values from data fields resident in a system file or database. The system file or database from which the values are extracted is referred to as the data source. Interim processing can occur and the data is then stored in another system file or database referred to as the target. Any data retrieval and reporting tool can be used as an extraction tool provided it has the ability to store the retrieved data in a form that can be used as input to a subsequent retrieval tool. There are software programs, however, referred to as ETL (Extraction, Transformation, and Loading) tools that are specifically designed for this purpose. They generally provide the ability to access a number of different types of databases and data file formats.

Data profiling is the process for examining and analyzing characteristics of data to ascertain or improve the level of its quality. The importance of data profiling is heightened in those instances where allegedly identical (or predominantly similar) data is intended to be extracted from multiple sources, then integrated and stored into a single target structure. A basic data profiling approach would include data column analysis where specific properties are measured for individual data elements, e.g., minimum, maximum, and average field lengths, or minimum, maximum and mean for numeric values, precision and scale for numeric values, data type, data format, the number of distinct values for the field, and the number of occurrences of null/empty values. Depending upon the type of field and the count of its distinct values, a complete list of values compared to known valid values should be produced also.

Data quality is the process of rectifying data defects and improving the accuracy, integrity, and understandability of data. This is preferably accomplished by identifying and correcting inconsistent or erroneous data in a source operational system or data warehouse architecture. However, if it is not cost feasible (or perhaps organizationally feasible) to implement data quality enhancements at those upstream stages, then it falls to the Actuarial Data Management unit to effect a reactive data improvement process at the department level.

4.3.3 Data Integration And Data Transformation

Data integration is the process of merging data from different (and sometimes very disparate) sources. Source systems that contain what is thought to be identical data elements can, in fact, prove to contain different formats, data types, and representations of information. For example, four different systems may contain a data field identified as STATE. In one system, the value stored in a data record could be "Alabama", in another, "AL", in another "01", in another, the field may be null. To consolidate information from multiple systems in a manner that either retains or introduces integrity to the data, the source data must be understood. Likewise, to establish data integrity in a target data structure, target data elements must comply with the business and technical requirements driving the integration task.

Effective data integration involves the application of business rules (systematic procedures) for "cleaning" data and deriving valid and as-accurate-as-possible versions of it. This is necessary to ensure that actuaries who will eventually use the data as input to their analyses need not qualify their results with excessive error margins. The subprocesses of converting data to different data types and formats, applying data cleansing techniques, consolidating varied representations of the same values, e.g., "01" vs. "AL", and deriving new data elements from others, can be referred to as data transformations.

Data transformations that are used to populate operational data structures can be very different from those used for actuarial data structures. In the cases of operational systems and data warehouses, data elements are normally static with respect to historical occurrences. That is, when changes to the definitions or valid values of data elements in these systems occur, the changes are made from a current point in time and only for <u>subsequent</u> periods thereafter. Redefinition and valid value descriptions are not enacted retroactively.

In the case of actuarial data structures, it is not uncommon, especially for user-defined fields, to recast data definitions and data element values for all time periods. In fact, such redefinitions (and repopulating of the data in the structure) may occur regularly, each year, quarter, or even month. To non-actuarial data managers, this practice seems illogical and mistakenly viewed as a violation of proper data management rules and practices. To an Actuarial Data Manager or an Actuarial Technician, however, it represents a critical added value that he or she brings to the data resources they deploy to their actuarial clients. Where an actuarial staff is dependent upon non-actuarial data management resources, it will likely be necessary to explain the basis for such dynamically changing fields to obtain the necessary views of the data. Unfortunately, even with sufficient explanation, the normal flow of data projects through IT areas may not allow timely turnaround of such requests. Apart from the need for specialized subject matter expertise and attentiveness to data quality matters, this issue of addressing dynamic data requirements in a timely manner is a leading reason why an actuarial area needs its own data management unit.

4.3.4 Data Loading

The process of loading data into a structure for retrieval by actuarial personnel varies according to the type of database or retrieval tool that is intended to be used. For example, interim processing programs can save relatively small data tables as Microsoft Access database files, or even as comma-separated value (CSV) files for later importation by actuarial analysts. Oracle files can be loaded using native structured-query language (SQL) commands or using a more robust tool specifically designed for mass loading of data such as Oracle *SQL*Loader* (pronounced Sequel Loader). As the name implies, any ETL (Extraction-Transformation-Loading) software tool incorporates loading functionality into its design. But if such a tool were not already procured for the purposes of data extraction and/or transformation, it is likely be cost prohibitive to purchase it exclusively for loading data into a database structure.

5. CASE STUDY DISCUSSION

5.1 An Introduction To GMAC Insurance

In order to demonstrate a real-life scenario of the evolution of actuarial data management (ADM) functions within a high-volume transaction processing environment (HVTPE), the authors would like to describe a bit of history taken from their collective experiences at GMAC Insurance. And to place their experiences into the context of an evolving business organization, a brief background of GMAC Insurance is provided.

GMAC Insurance traces its roots to 1925 when the General Exchange Insurance Corporation (GEIC) was founded as a subsidiary of the General Motors Acceptance Corporation (GMAC). GEIC was established to fulfill the insurance needs of GM dealers and their customers. Initially, the focus was on physical damage protection for automobiles. A year after its incorporation, GEIC became the largest writer of automobile physical damage insurance in the United States and Canada. In 1939, GMAC established an agency company named the Motors Insurance Corporation (MIC), for which GM dealers became licensed insurance agents. And in 1960, GEIC and MIC merged, retaining only the latter's name.

During the 1970s, Motors Insurance Corporation pioneered mechanical repair protection. This optional coverage provided financial protection to customers of GM dealers for certain automobile repairs and services that were outside the scope of the traditional vehicle warranty. Examples of these repairs and services included warranty-type repairs occurring beyond the months and miles provisions of the warranty, rental car reimbursement for multi-day repairs, towing of the mechanically-disabled vehicle to a dealership repair facility, and a waiver of the applicable deductible for in-warranty repairs. The popularity of the mechanical repair protection programs catapulted the business line to a prominent status among MIC's writings by the end of the 1980s. The rapid growth of the mechanical repair protection business made it necessary to continually seek improvement of the data capture, reporting, and analysis processes managed by the actuaries and support people assigned to those tasks.

During that same era, the company expanded its products and services to transform into a truly multi-line property/casualty underwriter. Throughout the 1990s to the present, as opportunities have arisen to better support and add value to its parent organizations, General Motors Corporation and GMAC, the Motors Insurance Corporation has teamed up with other organizations through acquisitions and specialty insurance startups to form the GMAC Insurance Group, an A-rated (Excellent), Top 30 insurance group according to A.M. Best, with combined insurance writings approaching $3 billion per year.[3] GMAC Insurance remains the leader in the mechanical repair protection business, offering coverage on new and used, GM and non-GM vehicles sold throughout the United States and Canada.

5.2 The Rudiments Of Actuarial End User Computing

As one would expect, just as the mechanical repair protection business has substantially evolved over the past 25+ years at GMAC Insurance, so too have the end user computing practices and capabilities evolved that support its actuarial functions. In the beginning, the tracking of premiums and losses was an elementary task, although only very basic hard copy reports containing summary information were available for that purpose. At a system level, the corresponding contracts were recorded in a manner similar to automobile physical damage policies. This was a reasonable method for coding the information at the time, since losses were typically limited in size, and quickly paid once submitted. In fact, the initial pricing reports were produced in a format identical to the organization's automobile physical damage summary reports.

By the early 1980s, the pricing and reserving analysis work relied on very detailed reports produced by programs resident on the same mainframe platform as the mechanical business processing system. The reports required multiple boxes of computer paper to print out the detail needed as input for actuarial analyses. The required data typically resided many pages apart, and the sheer volume of paper to be stored required an excessive amount of physical storage space. Summary levels of the data had been predetermined and programmed into the reports. Consequently, if a subset of experience was needed, summarization proceeded manually, flipping through the pages and separately recording each aggregate. If a deeper level of detail was required, the only recourse was to request a programming change to develop another paper report. A revised report would normally require a minimum of a few weeks to a few months depending upon the complexity of the request and the workload (or backlog of requests) in the Data Processing Department.

The next innovation was the use of microfiche. The same volume of data took up much less space. Multiple timeframes were easily stored in a "shoebox" file, whereas the paper reports had required a storage room. However, the same data manipulation issues remained. Information was not easy to summarize, and if an additional level of detail was required, programming changes were needed.

A new method of obtaining new levels of summary from the microfiche reports was needed. An actuarial support person experienced in compiler languages (*Fortran* and *PL/1*) as well as fourth generation languages (*Focus* and *Easytrieve*) gained access to the master files of the reporting system. After creating extract files for each major segment of the mechanical business, summary level reports became readily available on an ad hoc basis. As needed, new levels of summary could be created and printed with one or two days of notice to the technician. This approach radically reduced the turnaround time to aggregate data by new criteria. If additional detail was required, however, a special request to the Data Processing Department was still necessary to effect changes to the production reporting system.

As the 1980s elapsed, the advent of personal computing began changing the analysts' landscape at the company. The ability of the end user to manipulate and summarize larger and larger subsets of information allowed pricing analysts to consider managing more detailed views of the business. Database programs on a desktop PC meant multiple summaries of a common set of data could be produced and compared to one another. Spreadsheet programs allowed the analysts to do more than just summarize larger sets of information. It was now possible to adjust the data for known influences, and thereby ferret out a deeper level of understanding of similar yet distinct segments of the business.

For the actuary, the PC provided a locally controlled, adaptable, "real-time" tool for analyzing loss triangles, exhibiting the policy year emergence of premiums and losses. Classification plans could be more frequently reviewed. "What if" analyses could be completed within the pricing and reserving functions, without the additional time burden of external programming efforts.

Unfortunately, with the increased use of personal computing, there came an increased demand for access to the raw material of data analysis, viz. data. It became quickly apparent that the process of re-keying data from mainframe-generated reports into the PC environment was costly, time-consuming, slow, and rife with the potential for input errors. Clearly, better and more efficient ways to gain access to the data were needed.

The next approach to improved data entry was to find a way to move from paper and microfiche reports to electronic versions of the reports. An early attempt entailed electronically capturing the online "print files" corresponding to the hard copy data reports. This information was then parsed and downloaded into a format accessible to the PC tools available at the time.

On the upside, the information no longer needed to be re-keyed, and the information was more easily balanced back to the original reports. Also, summarization proceeded more quickly via PC database programs. The PC environment allowed the actuary to adapt the analysis to reflect changes in the pricing and business environment. Grouping related segments of business for common analysis was much easier to do.

On the downside, the electronic data was captured at a highly summarized level. No greater level of detail could be extracted from these reports, without returning to the original sources of data, i.e., the mainframe online files within the mechanical business processing system. Without an apparent alternative, the actuaries and pricing analysts continued to rely on data programming professionals to accomplish refinements to the existing reports. This reliance led to large gaps in time between data requests and the subsequent retrieval and analysis of results. In addition, the process was slowed down by the need for the actuaries and analysts to explain fairly technical data requirements to the programmers who were not acquainted with actuarial analysis processes.

5.3 Getting Access To All The Raw Data

In the early '90s, it became evident that increasingly complex analyses required increasingly detailed information to support the analyses. Rather than continue the process of programming ad hoc subsets of the online files, it was proposed that a comprehensive set of mainframe data files be constructed that would make all of the contract level information available for actuarial analysis. The files were intended to meet the primary requirements of the pricing function and would capture the history of mechanical business as far back as could be retrieved. Consequently, the file set was dubbed the "Pricing History Data Files" or Pricing History Files (PHF) for short. These data files were to provide a very detailed, inception-to-date snapshot of all premium and loss records at a vehicle level as of each month end.

When contemplating the design of the PHF and the mechanics of building them, two approaches were considered. One was to construct files containing incremental transactions from the inception of the contract for each vehicle. Each month a new set of incremental records would be created and inserted into the files. The inception-to-date view of the business could then be derived by aggregating the records as of a given evaluation date. The second approach was to create an aggregate inception-to-date record for each vehicle from an existing "snapshot" master file, then combine additional incremental experience into the records as each month elapsed. The first option would allow the analysts to scrutinize the data down to the monthly operational level of detail. The second would not provide that granularity, but it would provide a level of detail sufficient for advancing the current state of pricing and reserving analyses. Ultimately, the pricing analysts chose the second option, due to the substantial additional cost (in terms of programming resources, data storage overhead, and ongoing processing time) expected to accompany the preferred first option.

The PHF System was created and controlled at the department level. The in-house system engineer contracted to technically design and build the data files worked side by side with the pricing analysts and actuaries on a daily basis. The resulting PHF data files were constructed in the remote mainframe environment, but extracts were summarized and downloaded for use in the local PC environment. Additional pricing details, such as descriptions of encoded values, were added to the PHF data. In the end, the PHF System enabled the analysts to "slice and dice" premium and loss experience across multiple time frames, blocks of business, vehicle types, etc.

With the programming and data resources now under departmental control, data extractions and summarizations were accomplished in a timelier manner. The analysts were able to communicate directly with a dedicated programmer, and so over time it became easier for both to collaborate on describing, accessing, and refining data requirements. As time progressed, several of the analysts became proficient in programming retrievals of their own data from the PHF System.

5.4 The Standardization Of The Data Extractions

Over time, it was observed that while many different analyses were created from the PHF System, there were common traits to a significant subset of data requests. For example, while a group of reports might have included loss amounts and loss counts arrayed by time intervals, the aggregations were constantly changing. For one analysis, aggregation might have been by vehicle type, while for another analysis the aggregation might have been by level of coverage. It was not uncommon to find an analyst describing their data request as "just like the last one except…" The upshot of this common thread was a realization that by creating standardized extracts from the PHF System, multiple data requests could be handled at once. (In today's world, these data extracts might be called "data marts," rather than databases.) These databases were created in a standardized format, at predetermined time intervals. The actuaries and pricing analysts would then summarize the databases to render useful information for their respective data analyses.

The business data needed for actuarial analysis of the mechanical repair programs was now accessible in a timely fashion. But to ensure the ongoing sustainability of the new data preparation process, a new role to complement the role of the pricing analysts and actuaries was needed. This role would oversee the data preparation processes from the routine regeneration of the PHF, to the subsequent extraction, summarization, downloading, and balancing of the data. And the new role would prove to be an impetus towards the next step in end user computing within the company's actuarial area.

5.5 The Emergence Of The ADM Function

The Actuarial Data Management function emerged in the mid-1990s. In its role as overseer of the actuarial pricing data preparation process, the ADM unit became responsible for balancing the PHF System output, fulfilling standardized database requests, organizing the resultant data sets, and creating complex applications in response to new analytical data requests by the actuarial analysts. The PHF, the standardized databases, and the processes and programs surrounding them were placed under the control of the ADM unit, comprised of a business manager with IT management experience, an actuarial technician, and two technical contractors.

What made this process work well was the division of labor. The ADM team was able to focus on the data development and delivery processes. The pricing analysts and actuaries were free to focus on using the information contained in the data. Both groups benefited from working together under a common departmental structure. Potential changes to data requirements could be discussed, tested, and refined based on direct interaction between the ADM team and the analysts.

The ADM team also provided the necessary bridge to enhance data management functions for actuarial use. For example, rather than rely solely on a suite of PC-based tools for working with small databases and spreadsheets, the ADM team investigated the use of OLAP (Online Analytical Processing) tools for storage and delivery of information. As a result of their efforts, an OLAP tool was procured and installed on a local PC server. This allowed for the routine deployment of "data cubes" to the actuaries as part of the routine PHF data refresh process.

The OLAP tool provided the actuarial analysts with fast access to multidimensional data. Much of the summarization was pre-computed, so that OLAP data retrieval was nearly instantaneous. Analysts gained the ability to query and retrieve multiple views of requested information, summarized along various dimensions or "cuts" of the data. Because the OLAP data was centralized, very little computing power was required at the local (client) PC. Rather, the OLAP server performed this function. Some have characterized the OLAP environment as an "industrial-strength pivot table." Suffice it to say that the OLAP tool implemented at GMAC Insurance came to be regarded as all that and more.

5.6 The Implementation Of The Data Warehouse

By the late 1990s, several other departments within the company began to look for ways to extract meaningful data to meet their informational needs. The views of the business data that were needed to serve their purposes, however, were different than those of the mechanical business pricing function. Claims personnel sought detail at the vehicle repairer level and at the individual claim level. Marketing required sales summaries by region and distribution channel. Finance needed detail to support their aggregations of company and line of business financial reports.

It also became clear that company management desired a "single source" for mechanical data reporting, which could replace a number of special purpose reporting systems with their individual support requirements and escalating maintenance costs. The benefits of such a strategy would include increased quality and consistency of information across departments, a reduction in the long term maintenance of the data, and centralized management and access to the data. The "single source" would need to be a "production" system, i.e., a resource recognized as part of company data processes, not a departmental data system. The company was on the threshold of designing and building a data warehouse.

As the data warehouse initiative was formally launched in 1997, the Actuarial Department stepped forward to provide key business support. The Vice President & Chief Actuary served as Project Champion, and subject matter experts from both the Mechanical Pricing & Reserving function and the Actuarial Data Management function participated. Project team members from other departments within the company participated in requirements capturing sessions and served as subject matter experts for their respective functions with the company. The IT area managed the project as well as the vendor used to design and build the warehouse. The Mechanical Data Warehouse rolled out in phases with the final phase deployed to production status in mid-2000. The completed project resulted in an extremely detailed and useful corporate data warehouse for the mechanical business line, with information stored at the transactional level. Since that time the Actuarial Data Management unit has served as the company's maintainer of the Data Warehouse Data Dictionary, and performs detailed balancing procedures as part of the data warehouse monthly refresh process. It also performs the company's user acceptance testing of data stored in the warehouse when changes are made to upstream (source) systems.

5.7 The Migration From The PHF System To The Data Warehouse

The deployment of the Mechanical Data Warehouse provided the foundation for re-sourcing the data used by the actuaries from the PHF System to the new and single corporate source for mechanical business data. The eventual migration of the PHF applications and the re-sourcing of the actuarial databases, which transformed them into true data marts, was a development project far surpassing any single ADM group effort to that point in time. The knowledge, experience, and advanced skills of the unit's senior actuarial technician and its technical contractor were critical for achieving a successful transition. Their efforts combined with that of the other support technicians made it possible to complete the project while maintaining a demanding ADM production schedule and providing data warehouse UAT support for numerous corporate initiatives that ran concurrently.

6. CONCLUSION

To conclude this discussion paper, the authors would like to reiterate the goal of an Actuarial Data Management unit, and provide some helpful perspective when attempting to establish or advance the function within an insurance organization.

6.1 The Goal Of Actuarial Data Management

As first stated at the beginning of this paper, the goal of an Actuarial Data Management unit is to equip an actuarial staff with the data resources necessary to excel in the performance of its functions. The ability of the ADM team to achieve that goal in a high-volume transactional processing environment depends upon many factors, some that are under the unit's control, and some that they can only exert influence upon.

Factors under the control of the ADM unit include attitudes and behaviors conducive to high quality work performance and high quality data development processes. These are best expressed by an orientation towards ongoing incremental improvement of the technical processes, persistence at resolving issues and problems related to data quality and timeliness, and perseverance during periods of heightened difficulties and setbacks. Another important factor under the unit's control is adequate testing of both interim and final deliverables created to satisfy actuarial data resource requirements.

Important factors upon which influence can be exerted are the requirements gathering process, interactions with actuarial clients and the IT area, education and training in the tools used to perform ADM functions (including participation in software user groups), roles in user testing of corporate projects that impact ADM functions and supported data resources, and the visibility of end user computing capabilities within the business and IT sectors of the company.

In some organizations, the actual tools used to perform ADM activities can be selected or at least influenced. In others, the tools are prescribed by standards established by the IT area or by corporate policy. Likewise, the degree to which technology can be applied to business problems by business people (who are not IT professionals) is sometimes governed by the IT area or corporate policy. In that regard, business-led actuarial data management activities can sometimes be perceived to press upon and even push the limits of what a business function should undertake. However, when a technical solution is within the capabilities of an ADM function and its developmental scope is departmental, not corporate, the IT area should recognize that their responsibilities need not extend further than administering the underlying infrastructure and providing advisory support as requested to resolve issues.

6.2 Actuarial Data Management Now And In The Future

Actuarial Data Management is a necessary function in a high-volume transactional processing environment. It serves as a proactive, added-value conduit of business data and specialized technical support to the actuarial staff. The greatest value will be realized by actuarial departments that foster within their ADM function a balance of subject matter expertise in actuarial data requirements and advanced technical competence with information technology. Such a balance not only makes it possible for the ADM unit to meet its prime directive of providing analysis-ready data to its actuarial clients as needed, but also to serve as a bridge between the actuarial staff and the IT area, facilitating much of their interactions and expediting mutual interdepartmental objectives.

The challenges of establishing and maintaining an effective actuarial data management function today and in the future will certainly persist and in some respects escalate. This is evidenced in the limited availability of highly qualified actuarial technicians, the growing need to ensure procedural integrity and consistency in compliance with federal regulations such as the Sarbanes-Oxley Act of 2002, the perpetual release of software and hardware innovations that support data management activities, the commensurate withdrawal of vendor support of past generations of software and hardware, and the continuing evolution of actuarial data research and analysis techniques. These current issues are compelling changes in the ADM function simply to maintain the status quo. When taken into account with the perennial drivers of change associated with an insurance organization (revenue/income enhancement, customer acquisition/satisfaction/retention, loss/expense containment, and marketplace competition), there can be no expectation of respite on the part of ADM professionals in support of their actuarial clients.

Endnotes

[1] John F. Rockart and Lauren S. Flannery, "The Management of End User Computing," Communications of the ACM, Vol. 26, Issue 10 (October 1983), pp. 776-784, ACM Press, New York, NY.

[2] Ibid.

[3] "U.S. Ranking by Assets – Groups" and "U.S. Ranking by Net Premiums Written – Groups", Best's Aggregates & Averages, Property/Casualty, 2004 Edition, A.M. Best Company, Oldwick, New Jersey, pp. 654, 658.

General References

Building the Data Warehouse, Second Edition, W. H. Inmon, John Wiley & Sons, New York, NY, 1996.

Building, Using, and Managing the Data Warehouse, Ramon Barquin & Herb Edelstein, Ed., The Data Warehousing Institute Series, Prentice Hall, Upper Saddle River, New Jersey, 1997.

Planning and Designing the Data Warehouse, Ramon Barquin & Herb Edelstein, Ed., The Data Warehousing Institute Series, Prentice Hall, Upper Saddle River, New Jersey, 1997.

General Motors: Priorities and Focus – Yesterday, Today and Tomorrow, First Edition, General Motors University, College of Leadership, General Motors Corporation, Detroit, Michigan, November 2000.

Abbreviations and Notations

ADM, Actuarial Data Management
CSV, Comma Separated Value
DP, Data Processing
ETL, Extraction-Transformation-Loading
GEIC, General Exchange Insurance Corporation
GMAC, General Motors Acceptance Corporation
HVTPE, High-Volume Transactional Processing Environment
IS, Information Systems
IT, Information Technology
LAN, Local Area Network
MIC, Motors Insurance Corporation
OLAP, Online Analytical Processing
PHF, Pricing History Files
SQL, Structured Query Language
UAT, User Acceptance Testing
WAN, Wide Area Network

Biographies of Authors

Joseph Strube is Sr. Manager of Actuarial Data Management at Motors Insurance Corporation, a member of the GMAC Insurance Group, located in Southfield, Michigan. His team develops and maintains analytical databases for the Home Office actuarial staff and oversees the data warehouse for the organization's largest business line. A graduate of the University of Michigan, Mr. Strube's career in the property/casualty insurance field has spanned 30 years at multiple carriers and includes assignments in the Actuarial, Financial, and Information Technology areas. Mr. Strube draws upon his management and technician experiences in the IT area as well as his current management role over actuarial support functions.

Dr. Bryant Russell is a Sr. Team Leader in the Actuarial Department of Motors Insurance Corporation, a member of the GMAC Insurance Group, located in Southfield, Michigan. He has had extensive experience in pricing and reserving for long-term automobile extended service contracts. He was one of the actuarial team members engaged in creating a corporate actuarial database for pricing and reserving of such service contracts. Dr. Russell later served as actuarial subject matter expert during the development of the organization's data warehouse. He achieved his A.C.A.S. designation in May 2000. Prior to entering the actuarial profession, Dr. Russell completed his Ph.D. in Mathematics at the University of Michigan.